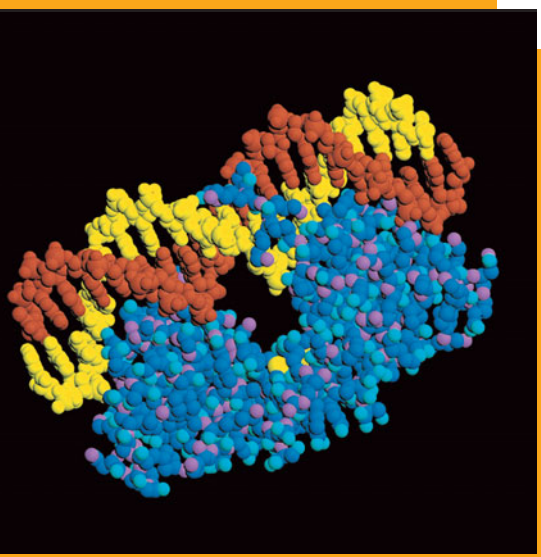
17 Regulation of Gene Expression in Bacteria and Bacteriophages



lac operon repressor protein binding to DNA.

Key Questions

- How is gene expression regulated in bacteria?
- How is gene expression regulated in bacteriophages?

iActivity

ONE OF THE BEST STRATEGIES FOR AN organism's survival is to be able to adapt quickly to changes in its environment. In fact, a fundamental property of living cells is their ability to turn their genes on and off in response to extracellular signals. This control of gene expression makes it possible for cells to produce specific kinds of proteins when and where they are needed. In this chapter, you will learn some of the ways in which gene expression in microorganisms is regulated. Then, in the iActivity, you can investigate how mutations affect the process of regulation in *E. coli.*

Most bacteria are free-living organisms that grow by increasing in mass and then divide by binary fission. Growth and division are controlled by genes, the expression of which must be regulated appropriately. Genes whose activity is controlled in response to the needs of a cell or organism are called **regulated genes.** All organisms also have a large number of genes whose products are essential to the normal functioning of a growing and dividing cell, no matter what the conditions are. These genes are always active in growing cells and are known as **constitutive genes** or *housekeeping genes*; examples include genes that code for the enzymes needed for protein synthesis and glucose metabolism. Note that all genes are regulated on some level. If normal cell function is impaired for some reason, the expression of all genes, including constitutive genes, is reduced by regulatory mechanisms. Thus, the distinction between regulated and constitutive genes is somewhat arbitrary.

The goal of this chapter is to learn about some of the mechanisms by which gene expression is regulated in bacteria and bacteriophages. Significantly, genes which encode proteins that work together in the cell typically are organized into operons; that is, the genes are adjacent to each other and are transcribed together onto a *polycistronic mRNA*, so called because it contains the information from more than one gene. (Here, the word *cistron* is used synonymously with *gene*.) Regulation of the synthesis of this mRNA depends on interactions between regulatory proteins and regulatory sequences that are next to the gene array. Such studies of bacterial and bacteriophage gene regulation also have provided important insights into how genes are regulated in higher organisms, including humans. Of course, much remains to be done to understand completely the regulation of gene expression in bacteria. The 4.6-megabase (4.6×10^6 bp) genome of *E. coli,* for example, has 4,288 protein-coding genes according to the genomic sequence. Genomics researchers are able to say something about the function of approximately 80% of those genes, but much remains to be determined about their complete functions and how they are regulated. And, the functions of the remaining approximately 20% of the genes remain unknown. This chapter's Focus on Genomics box describes a computer model for regulation of gene expression in one prokaryote, *Halobacterium salinum*.

Focus on Genomics
Models of Gene Expression

The combination of genomics, transcriptomics, proteomics, and DNA microarrays has allowed scientists to begin building models that predict how a cell will respond to environmental change. These models can predict which genes will be expressed under certain conditions, and which DNA-binding proteins will regulate the expression of these genes. In one such study, bioinformatics was used to predict the likely function, where possible, for each of the 2,400 proteins encoded in the complete genomic sequence of the archaean *Halobacterium salinum*. The investigators then performed an extensive series of DNA microarray experiments examining the expression of all 2,400 genes under specific environmental conditions. For instance, they compared gene expression in cells grown in the presence of high levels of nickel with that of cells grown in environments without nickel. They also used DNA microarrays to test the functions of specific genes. For example, they compared gene expression in wild-type cells with cells containing a mutation that prevented the production of the transcription factor TFBf. Using this kind of data, the researchers were able to build a computer model that accurately predicted the transcriptional response of the cell to new environmental or genetic challenges. When they tested their model against real-life experiments using new combinations of environmental or genetic changes, they found that the model accurately predicted the response of 80% of the expressed genes in the cell. As a result, they were able to organize the genes into *biclusters*, or sets of genes that respond in the same manner to a series of environmental and/or genetic changes, whose expression is likely regulated by the same cellular proteins. For instance, transcription of the 34 genes in bicluster 66 is regulated by the environmental factors oxygen and light and the transcription factors TFBf and Cspd1.

What can we learn from these models? We can extrapolate possible functions for genes that have no bioinformatics clues. For instance, bioinformatics told the investigators nothing about the function of the protein encoded by gene VNG1459H. This gene was placed in a bicluster with genes known to encode proteins that help the cell respond to light. This suggested a role for VNG1459H in light response, and the researchers were then able to find this protein localized to a region of the cell involved in sensing light.

We can also use this sort of analysis to modify microorganisms to help us with certain chemical or environmental processes. For example, if this approach is used to define biclusters in bacteria that degrade environmental toxins, we could then engineer bacteria to use these biclusters more effectively.

The *lac* Operon of *E. coli*

When gene expression is turned on in a bacterium by adding a substance (such as lactose) to the medium, the genes involved are said to be *inducible*. The regulatory substance that brings about this gene induction is called an **inducer,** and the phenomenon of producing a gene product in response to an inducer is called **induction.** The inducer is an example of a class of small molecules, called **effectors** or **effector molecules,** that help control the expression of many regulated genes. An inducible gene is transcribed in response to a regulatory event occurring at a specific regulatory DNA sequence adjacent to or near the protein-coding sequence (Figure 17.1). The regulatory event typically involves an inducer and a regulatory protein; and when it occurs, RNA polymerase initiates transcription at the promoter (usually upstream of the regulatory sequence). The gene is turned on, mRNA is made, and the protein encoded by the gene is produced. The regulatory sequence itself does not code for any product. As an example of such gene regulation, let us examine regulation of the genes of the *E. coli lac* operon, an **inducible operon.**

Lactose as a Carbon Source for *E. coli*

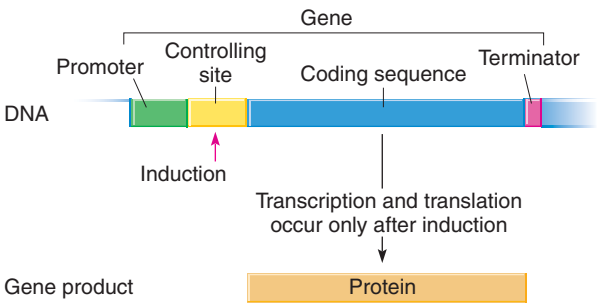
E. coli can grow in a simple medium containing salts (including a nitrogen source) and a carbon source such as glucose. The energy for biochemical reactions in the cell comes from glucose metabolism. The enzymes required for glucose metabolism are coded for by constitutive genes. If lactose is provided to *E. coli* as a carbon source instead of glucose, a number of enzymes that are required to metabolize lactose are rapidly synthesized. (A similar series of events, each involving a sugar-specific set of enzymes, is triggered by other sugars as well.) The enzymes are synthesized because the genes that code for them become actively transcribed in the presence of the sugar; the same genes are inactive if the sugar is absent. In other words, the genes are regulated genes whose products are needed only under certain conditions.

Lactose is a disaccharide consisting of the monosaccharides D-galactose and D-glucose. When lactose is present as the sole carbon source in the growth medium, three proteins are synthesized:

1. β-*Galactosidase.* This enzyme breaks down lactose into galactose and glucose, as well as catalyzing the

Figure 17.1

General organization of an inducible gene.



Inducible genes are expressed only in the absence of a repressor and/or presence of an effector/inducer molecule.

isomerization (conversion to a different form) of lactose to *allolactose*, a compound that is important in regulating expression of the *lac* operon (Figure 17.2). (In the cell, the galactose is converted to glucose by enzymes encoded by a gene system specific to galactose catabolism. The glucose is then utilized by constitutively produced enzymes.)

2. *Lactose permease* (also called *M protein*). This protein, found in the *E. coli* cytoplasmic membrane, actively transports lactose into the cell.

3. β-*Galactoside transacetylase.* This enzyme transfers an acetyl group from acetyl-CoA to β-galactosides. The function of this enzyme in the *lac* operon is not understood.

In wild-type *E. coli* growing in a medium containing glucose, only a low concentration of each of these three proteins is produced. For example, only an average of three molecules of β-galactosidase is present in the cell under these conditions. In the presence of lactose, and the absence of glucose, the amount of each enzyme increases coordinately (simultaneously) about a thousandfold (e.g., to about 3,000 molecules of β-galactosidase), because the three essentially inactive genes are now actively transcribed. The process is called **coordinate induction.** Allolactose, not lactose, is the inducer molecule directly responsible for the increased production of the three enzymes (see Figure 17.2). Furthermore, the mRNAs for the enzymes have a short half-life, so the transcripts must be made continually in order for the enzymes to be produced. When lactose is no longer present, transcription of the three genes is stopped and any mRNAs already present are broken down, so no more of these proteins are made. Existing proteins are degraded and diluted out by cell growth and division.

Figure 17.2

Reactions catalyzed by the enzyme β-galactosidase. Lactose brought into the cell by the permease is converted to glucose and galactose *(top)* or to allolactose *(bottom)*, the true inducer for the lactose operon of *E. coli*.

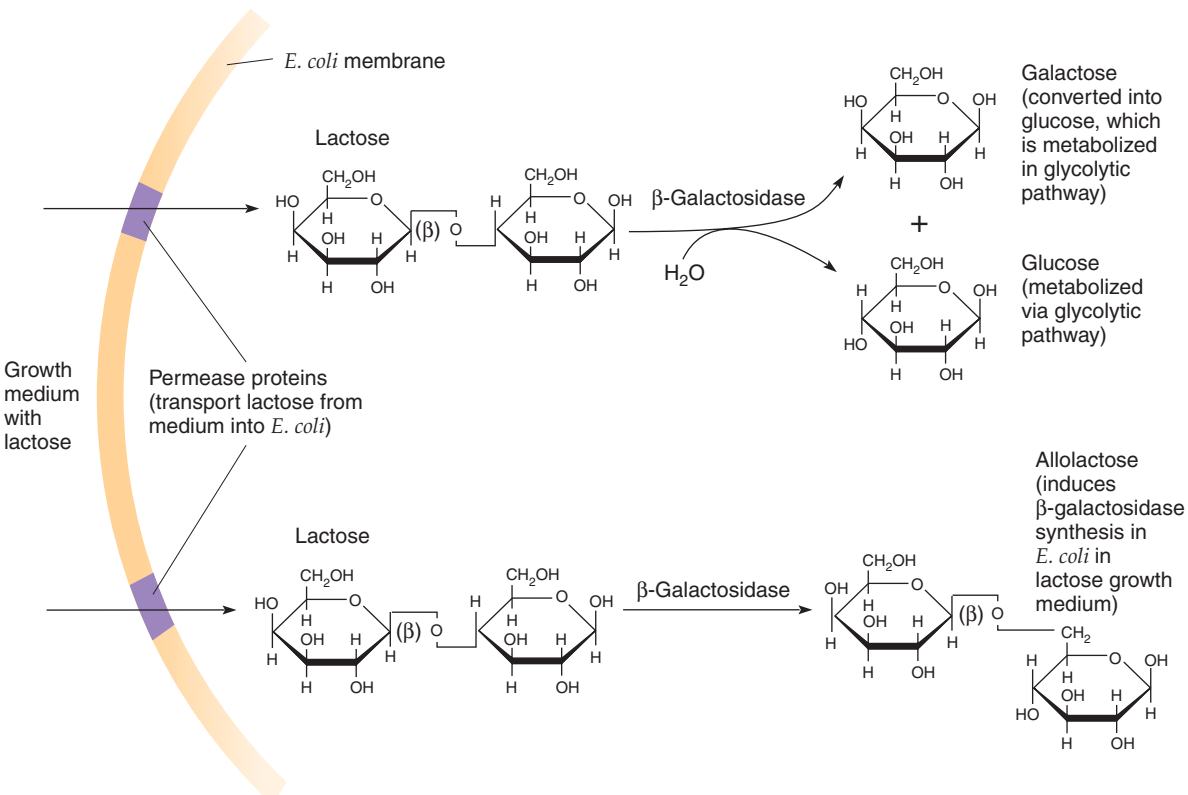
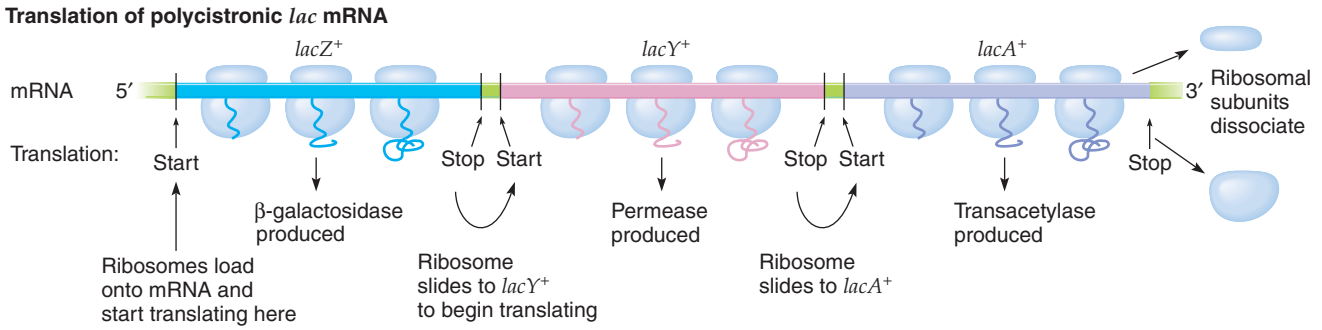

E. coli membrane
Lactose
β-Galactosidase
H2O
Galactose (converted into glucose, which is metabolized in glycolytic pathway)
+
Glucose (metabolized via glycolytic pathway)
Growth medium with lactose
Permease proteins (transport lactose from medium into E. coli)
Lactose
β-Galactosidase
Allolactose (induces β-galactosidase synthesis in E. coli in lactose growth medium)

Figure 17.3

Structure of the polycistronic mRNA encoded by the three clustered *lac* utilization genes in *E. coli* and its translation to produce β-galactosidase, permease, and transacetylase.



Experimental Evidence for the Regulation of *lac* Genes

Our basic understanding of the organization of the genes, the regulatory sequences involved in lactose utilization, and the control of expression of the *lac* genes of *E. coli* came largely from the genetic experiments of François Jacob and Jacques Monod, for which they shared (along with Andre Lwoff, for his work on the genetic control of virus synthesis) the 1965 Nobel Prize in Physiology or Medicine. Let us summarize their experiments.

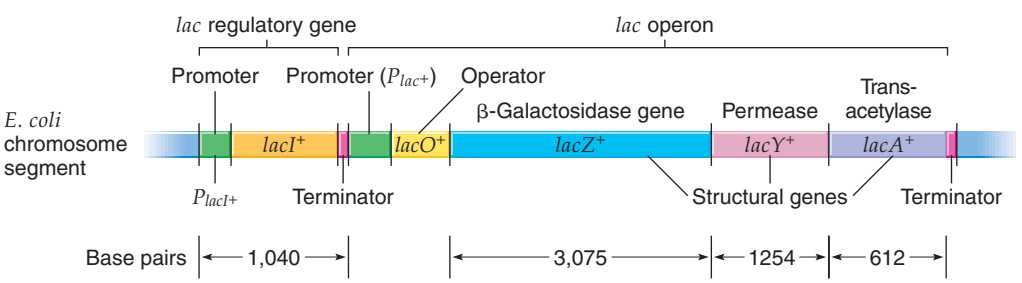
Animation
Regulation of Expression of the *lac* Operon Genes

Mutations in the Protein-Coding Genes. Mutations in the protein-coding (structural) genes were obtained after treatment of cells with chemical mutagens. The mutations were characterized to determine which enzyme activity was affected. The β-galactosidase gene was named *lacZ*, *the permease gene lacY*, and the transacetylase gene *lacA*. The *lacZ*$^-$, *lacY*$^-$, and *lacA*$^-$ mutations obtained were used to map the locations of the three genes via standard mapping experiments (see Chapter 15). The experiments showed that the three genes are tightly linked in the order *lacZ–lacY–lacA*. The three clustered genes were shown to be transcribed onto a single mRNA molecule—called a **polycistronic mRNA**—rather than onto three separate mRNAs (Figure 17.3). That is, RNA polymerase initiates transcription at a single promoter, and a polycistronic mRNA is synthesized with the gene transcripts in the order 5′-*lacZ*$^+$–*lacY*$^+$–*lacA*$^+$–3′. In translation, a ribosome loads onto the polycistronic mRNA at the 5′ end and synthesizes β-galactosidase; it then reinitiates translation at the permease sequence and synthesizes permease, reinitiates at the transacetylase sequence and synthesizes transacetylase, and finally dissociates from the mRNA.

Mutations Affecting the Regulation of Gene Expression. In wild-type *E. coli*, the three gene products are induced coordinately when lactose is present. Jacob and Monod isolated mutants in which all gene products of the operon were synthesized *constitutively*; that is, they were synthesized regardless of whether the inducer was present. The researchers hypothesized that the mutations were regulatory mutations that affected the normal mechanisms controlling the expression of the structural genes for the enzymes. They identified two classes of constitutive mutations: one class mapped to a small region just upstream of the *lacZ* gene they called the **operator** (*lacO*), and the other class mapped to a gene upstream of the operator they called the *lacI* gene or Lac **repressor gene.** Figure 17.4 depicts the organization of the *lac* structural gene cluster and the associated regulatory sequences. The promoter, operator, and the three structural genes constitute the *lac* operon.

Figure 17.4

Organization of the *lac* genes of *E. coli* and the associated regulatory elements: the operator, promoter, and regulatory gene. The promoter, operator, and three adjacent *lac* genes together constitute the *lac* operon.



Operator Mutations. The mutations of the operator were called operator-constitutive, or *lacO*^{c}, mutations. Through the use of partial diploid strains (*F′* strains in which a few chromosomal genes on an extrachromosomal genetic element called the *F* factor are introduced into a bacterial cell; see Figure 15.6, p. 434), Jacob and Monod were able to define better the role of the operator in regulating expression of the *lac* genes. One such partial diploid was $\frac{F'\ lacO^+\ lacZ^-\ lacY^+}{lacO^c\ lacZ^+\ lacY^-}$. (Both gene sets have a normal promoter, and the *lacA* gene is omitted because it is not important to our discussion.)

One *lac* region in the partial diploid has a normal operator (*lacO*$^+$), a mutated β-galactosidase gene (*lacZ*$^-$), and a normal permease gene (*lacY*$^+$). The other *lac* region has a constitutive operator mutation (*lacO*^{c}), a normal β-galactosidase gene (*lacZ*$^+$), and a mutated permease gene (*lacY*$^-$). This partial diploid was tested for production of β-galactosidase (from the *lacZ*$^+$ gene) and of permease (from the *lacY*$^+$ gene), both in the presence and the absence of the inducer.

Jacob and Monod found that active β-galactosidase is synthesized in the *absence* of the inducer and that permease is synthesized, but is inactive because of the mutation. Only when lactose is added to the culture and the allolactose inducer is produced is active permease synthesized. That is, the *lacZ*$^+$ gene (which is on the same DNA molecule as *lacO*^{c}) is *constitutively expressed* (meaning that the gene is active in the presence or absence of inducer), whereas the *lacY*$^+$ gene is under normal inducible control: it is inactive in the absence of inducer and active in the presence of inducer. In other words, a *lacO*^{c} *mutation affects only the genes downstream from it on the same DNA molecule*. Similarly, the *lacO*$^+$ region controls only *lac* structural genes adjacent to it and has no effect on the genes on the other DNA molecule. A gene or DNA sequence that controls only genes located on the same, contiguous piece of DNA is said to be ***cis*-dominant.** The *lacO*^{c} mutation is *cis*-dominant because the defect affects the adjacent genes only and cannot be overcome by a normal *lacO*$^+$ region elsewhere in the cell. In other words, the operator must not encode a diffusible product. If it did, then in the *lacO*$^+$/*lacO*^{c} diploid state, one would have controlled all the lactose utilization genes regardless of their location.

lacI Gene Regulatory Mutations. The second class of *lac* constitutive mutants defined the *lacI* gene. That is, *lacI*$^-$ mutants in a haploid cell have a constitutive phenotype. Again, the use of partial diploid strains illuminated the normal function of the gene.

The partial diploid here is $\frac{lacI^+\ lacO^+\ lacZ^-\ lacY^+}{lacI^-\ lacO^+\ lacZ^+\ lacY^-}$; both gene sets have normal operators and normal promoters. In the absence of the inducer, no β-galactosidase or permease was produced; both were synthesized in the presence of the inducer. In other words, the expression of both operons was inducible. This means that the *lacI*$^+$ gene in the cell can overcome the defect of the *lacI*$^-$ mutation. Because the two *lacI* genes are located on different DNA molecules (that is, they are in a *trans* configuration), the *lacI*$^+$ gene is said to be ***trans*-dominant** to the *lacI*$^-$ gene.

Because the *lacI*$^+$ gene controls the genes on the other DNA molecule, Jacob and Monod proposed that the *lacI*$^+$ gene is a repressor gene that encodes a **repressor** molecule, the Lac repressor. No functional repressor molecules are produced in *lacI*$^-$ mutants. Thus, in a haploid *lacI*$^-$ bacterial strain, the *lac* operon is constitutive. In a partial diploid with both a *lacI*$^+$ and a *lacI*$^-$, however, the functional Lac repressor molecules produced by the *lacI*$^+$ gene control the expression of both *lac* operons in the cell, making both operons inducible.

Promoter Mutations. The promoter for the structural genes (located at the *lacZ* end of the cluster of *lac* genes; see Figure 17.4) is also affected by mutations. Promoter mutants (P_{lac-}) affect all three structural genes. Even in the presence of inducer, the lactose utilization enzymes are not made or are made only at very low rates. Since the promoter is the recognition sequence for RNA polymerase and does not code for any product, the effect of a *P* mutation is confined to the genes that it controls on the same DNA strand. The P_{lac-} mutations are another example of *cis*-dominant mutations.

Jacob and Monod's Operon Model for the Regulation of *lac* Genes

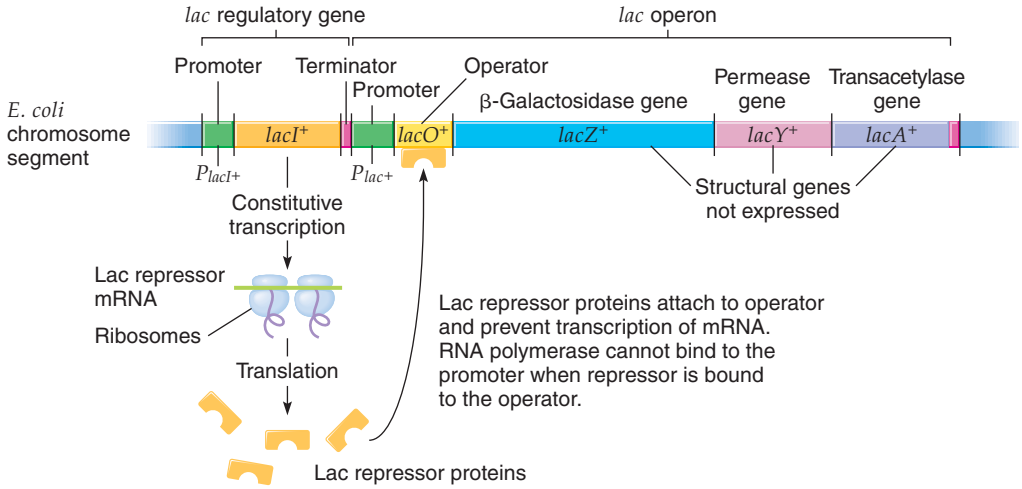
On the basis of their results, Jacob and Monod proposed their now-classic *operon model*. By definition, an **operon** is a *cluster of genes, the expressions of which are regulated together by operator–repressor protein interactions, plus the operator region itself and the promoter.* The promoter was not part of Jacob and Monod's original model; its existence was demonstrated in later studies. The order of the controlling elements and genes in the *lac* operon is promoter–operator–*lacZ*–*lacY*–*lacA,* and the regulatory gene *lacI* is located close to the structural genes, just upstream of the promoter (see Figure 17.4). The *lacI* gene has its own promoter and terminator and encodes the Lac repressor.

The description that follows of the Jacob–Monod model for regulation of the *lac* operon has been embellished with up-to-date molecular information. Figure 17.5 depicts the state of the *lac* operon in wild-type *E. coli* growing in the absence of lactose. The repressor gene (*lacI*$^+$) is transcribed constitutively, and the translation of its mRNA produces a 360-amino acid polypeptide. Four of these polypeptides associate together to form a tetramer, the functional Lac repressor protein (Figure 17.6). The promoter for the *lacI* gene is weak, so few repressor molecules are found in the cell.

The Lac repressor binds to the operator (*lacO*$^+$). The DNA sequence covered by the repressor protein overlaps

Figure 17.5

Functional state of the *lac* operon in wild-type *E. coli* growing in the absence of lactose.



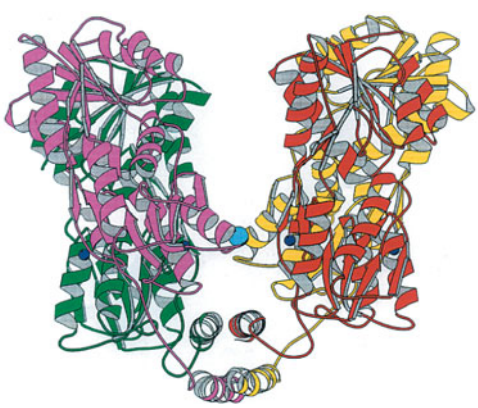
the DNA sequence recognized by RNA polymerase. Therefore, when the repressor is bound to the operator, RNA polymerase cannot bind to the promoter and transcription cannot occur—because of this, the *lac* operon is said to be under *negative control*. The low level of gene transcription that produces a few molecules of the enzymes, even in the absence of the inducer, occurs because repressors do not just bind and stay; they bind and dissociate. In the split second after one repressor unbinds and before another binds, an RNA polymerase could initiate transcription of the operon, even in the absence of the inducer. This "leaky" expression generates a few molecules of the three enzymes encoded by the *lac* operon, which is necessary to allow the initial transportation of lactose into the cell, and the initial conversion of lactose to allolactose when lactose is first added.

When wild-type *E. coli* grows in the presence of lactose as the sole carbon source (Figure 17.7), some lactose is converted by β-galactosidase into allolactose (see Figure 17.2). Allolactose binds to the Lac repressor and changes its shape; the shape change is called an *allosteric shift*. As a result, the repressor loses its affinity for the *lac* operator and dissociates from the site. Free repressor proteins are also altered by binding to allolactose so that they cannot bind to the operator. In this way, allolactose induces production of the *lac* operon-encoded enzymes.

With no Lac repressor bound to the operator, RNA polymerase initiates synthesis of a single polycistronic mRNA molecule for the *lacZ*$^+$, *lacY*$^+$, and *lacA*$^+$ genes. The polycistronic mRNA for the *lac* operon is translated by a string of ribosomes to produce the three enzymes specified by the operon. This efficient mechanism ensures the coordinate production of proteins of related function.

Figure 17.6

Molecular model of the *lac* repressor tetramer. The four monomers are colored green, violet, red, and yellow.



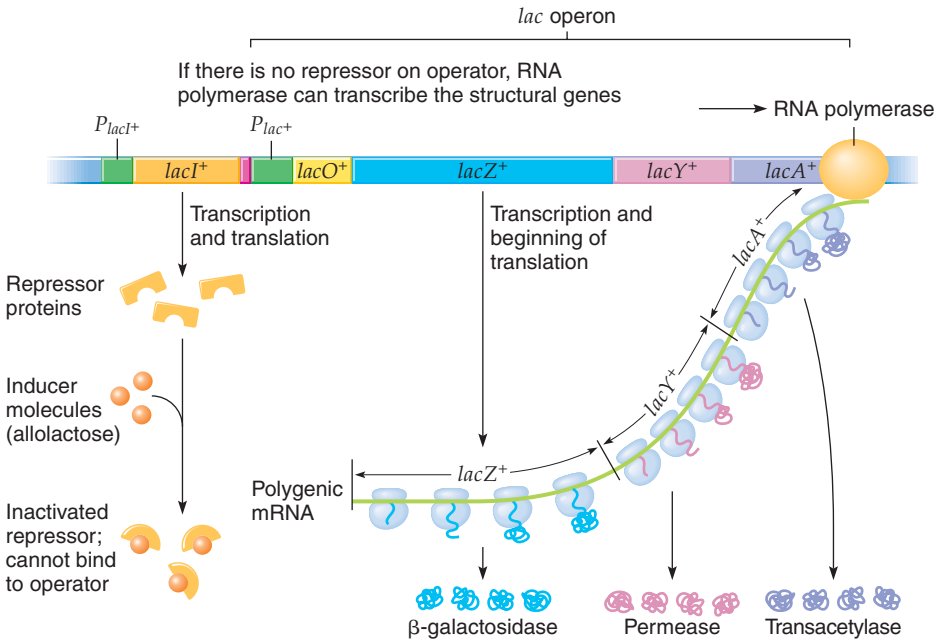
Effect of *lacO*^{c} Mutations. The *lacO*^{c} mutations lead to constitutive expression of the *lac* operon genes and are *cis*-dominant to *lacO*$^+$ (Figure 17.8). This is because base-pair alterations of the operator DNA sequence make it unrecognizable to the repressor protein. Because the repressor cannot bind, the structural genes physically linked to the *lacO*^{c} mutation become constitutively expressed.

Effects of *lacI* Gene Mutations. The *lacI* mutations map within the coding region of the repressor gene and result in changes to amino acids in the repressor polypeptide. The Lac repressor's shape consequently is changed or the translation in prematurely terminated, and it can neither recognize nor bind to the operator. As a consequence, in a haploid strain, RNA polymerase is not blocked from binding to the promoter and transcription cannot be prevented, even in the absence of lactose. As a result, constitutive expression of the *lac* operon occurs (Figure 17.9a).

The dominance of the *lacI*$^+$ (wild-type) gene over *lacI*$^-$ mutants is illustrated for the partial diploid $\frac{lacI^+\ lacO^+\ lacZ^-\ lacY^+}{lacI^-\ lacO^+\ lacZ^+\ lacY^-}$ described earlier. In the absence

Figure 17.7

Functional state of the *lac* operon in wild-type *E. coli* growing in the presence of lactose as the sole carbon source.



of the inducer (Figure 17.9b), the defective *lacI*$^-$ repressor cannot bind to either normal operator (*lacO*$^+$) in the cell. But sufficient normal Lac repressors, produced from the *lacI*$^+$ gene, are present, and they bind to the two operators and block transcription of both operons. When the inducer is present (Figure 17.9c), the wild-type repressors are inactivated, so both operons are transcribed. One produces a defective β-galactosidase and a normal permease, and the other produces a normal β-galactosidase and a defective permease; between them, active β-galactosidase and permease are produced. Thus, in *lacI*$^+$/*lacI*$^-$ partial diploids, both operons present in the cell are under inducible control.

Other classes of *lacI* gene mutants have been identified since Jacob and Monod studied the *lacI*$^-$ class of mutants. One of these classes, the *lacI*^{S} (*superrepressor*) mutants, shows no production of *lac* enzymes in the presence or absence of lactose. In partial diploids with a *lacI*$^+$/*lacI*^{S} genotype, the *lacI*^{S} allele is *trans*-dominant, affecting both operon copies (Figure 17.10). In this situation, the mutant repressor gene produces a superrepressor protein that can bind to the operator, but cannot recognize the inducer allolactose. Therefore, the mutant superrepressors bind to the operators even in the presence of the inducer, and the operons can never be transcribed. The presence of normal repressors in the cell has no effect, because, once a *lacI*^{S} repressor is on the operator, the repressor cannot be induced to fall off. Cells with a *lacI*^{S} mutation cannot use lactose as a carbon source.

A third type of repressor gene mutation is the *lacI*$^{-d}$ (*dominance*) class. In haploid cells, the *lacI*$^{-d}$ mutants have a constitutive phenotype like the other *lacI*$^-$ mutants; the *lac* enzymes are made in the presence or absence of lactose. Unlike the *lacI*$^-$ mutations, the *lacI*$^{-d}$ mutations are *trans*-dominant to *lacI*$^+$ in *lacI*$^{-d}$/*lacI*$^+$ partial diploids, so *lac* enzymes are produced constitutively even in the presence of the normal repressor.

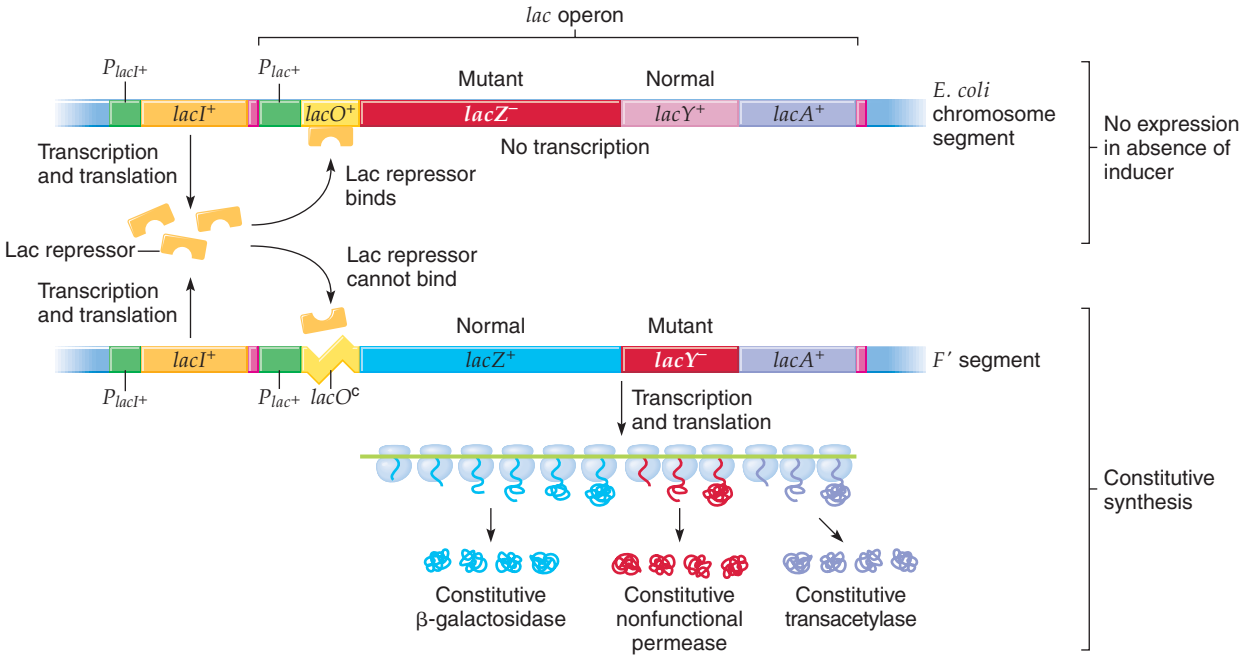
The dominance of *lacI*$^{-d}$ mutants is explained as follows: The Lac repressor protein is a tetramer consisting of four identical polypeptides. In *lacI*$^{-d}$ mutants, the repressor subunits do not combine normally, so no functional repressor tetramer is formed and no operator-specific binding is possible. The *lacI*$^{-d}$/*lacI*$^+$ diploids have a mixture of normal and mutant polypeptides, which combine randomly to form repressor tetramers. There are only about a dozen repressor molecules in the cell. The presence of one or more defective polypeptide subunits in the repressor tetramer is enough to block normal binding to the operator, so there is a good chance that no normal repressor proteins will be produced, because there are so few molecules per cell. As a result of the absence or near absence of complete, functional repressors, a constitutive enzyme phenotype results.

Finally, some mutations in the repressor gene promoter affect the expression of the repressor gene. Earlier, we indicated that the extent of transcription of a gene is a function of the affinity of that gene's promoter for RNA polymerase molecules. Since few repressor molecules are synthesized in wild-type *E. coli* cells, the repressor gene promoter must be of low affinity (i.e., it is a weak promoter). Base-pair mutations have been found that decrease and that increase transcription rates. For example,

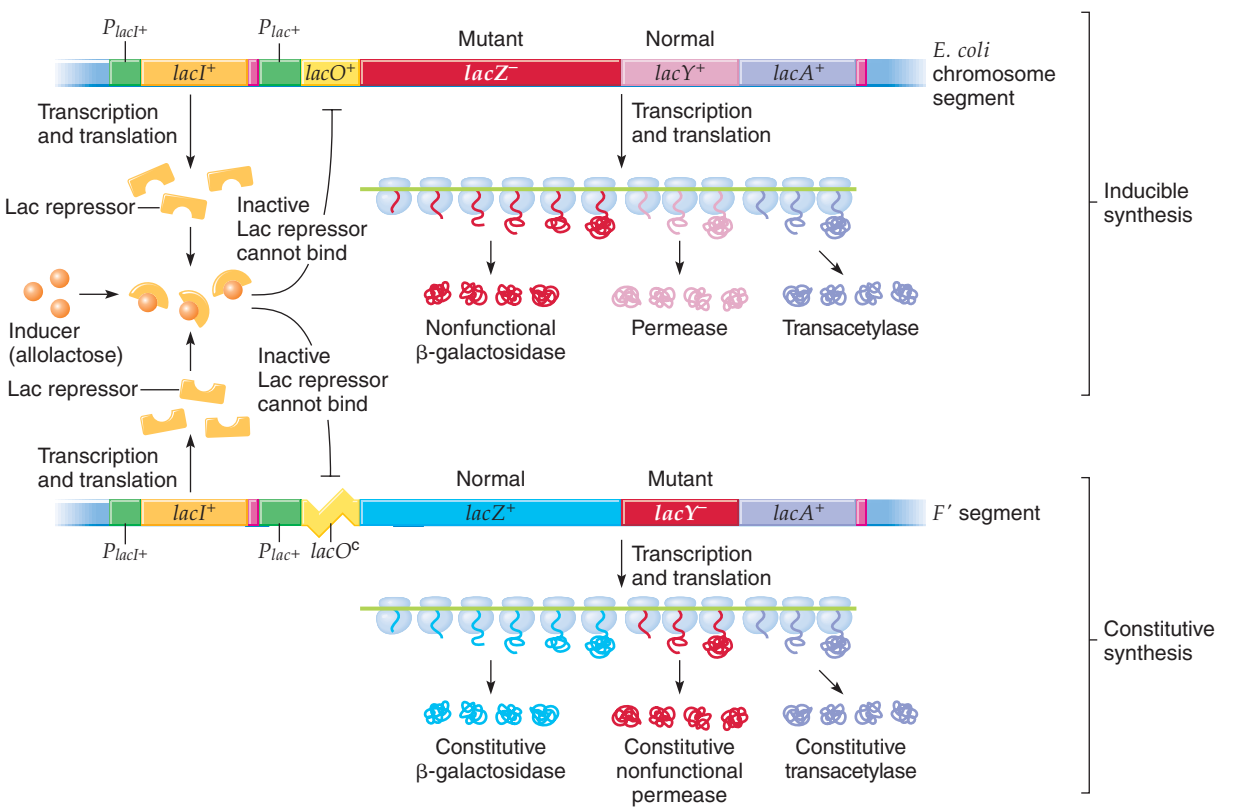
Figure 17.8

***Cis*-dominant effect of *lacO*^{c} mutation in a *lacI*$^+$ *lacO*$^+$ *lacZ*$^-$ *lacY*$^+$/*lacI*$^+$ *lacO*^{c} *lacZ*$^+$ *lacY*$^-$ partial diploid strain of *E. coli*.** (The *lacZ*$^-$ asqd *lacY*$^-$ mutations are missense mutations.)

a) Partial diploid in the absence of inducer. The *lacO*$^+$ operon is turned off, whereas the *lacO*^{c} operon produces functional β-galactosidase from the *lacZ*$^+$ gene and nonfunctional permease molecules from the *lacY*$^-$ gene with a missense mutation.



b) Partial diploid in the presence of inducer. The *lacO*$^+$ operon is turned on and produces nonfunctional β-galactosidase from the *lacZ*$^-$ gene and functional permease from the *lacY*$^+$ gene. The constitutive *lacO*^{c} operon produces functional β-galactosidase from the *lacZ*$^+$ gene and nonfunctional permease from the *lacY*$^-$ gene. Between the two operons, functional β-galactosidase and permease are produced.



lacI^{Q} and *lacI*^{SQ} mutants (where Q stands for "quantity" and SQ stands for "super quantity") result in an increase in the rate of transcription of the repressor gene, with the *lacI*^{SQ} giving the greater increase. These mutants were useful historically because they produce large numbers of repressor molecules, which facilitated their isolation and purification and, consequently, the determination of the amino acid sequence of the repressor polypeptide. Because *lacI*^{Q} and *lacI*^{SQ} mutants produce more Lac repressor molecules than the wild type produces, these mutants reduce the efficiency of induction of the *lac* operon. Note that *lacI*^{Q} and *lacI*^{SQ} can be induced at high lactose concentrations.

The mutants of the *lacI* gene point out the three different recognition interactions of the Lac repressor: (1) binding of the repressor to the operator region; (2) binding of the inducer to the repressor; and (3) binding of individual repressor polypeptides to each other to form the active repressor tetramer.

Positive Control of the *lac* Operon

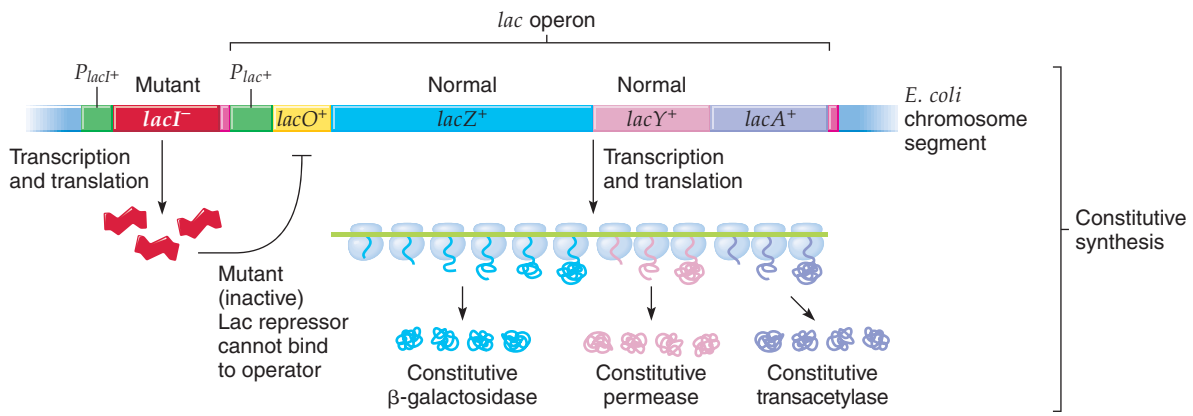
The Lac repressor protein exerts a negative effect on the expression of the *lac* operon by blocking RNA polymerase's binding to the promoter if the inducer is absent. Several years after Jacob and Monod proposed their operon model, researchers found a positive control system that also regulates the *lac* operon—a system that functions to turn on the expression of

animation
Positive Control of the *lac* Operon

Figure 17.9

Effects of a *lacI*$^{-}$ mutation in a *lacI*$^{-}$ *lacO*$^{+}$ *lacZ*$^{+}$ *lacY*$^{+}$ haploid cell, and in a *lacI*$^{+}$ *lacO*$^{+}$ *lacZ*$^{-}$ *lacY*$^{+}$/*lacI*$^{-}$ *lacO*$^{+}$ *lacZ*$^{+}$ *lacY*$^{-}$ partial diploid strain of *E. coli*. (The *lacZ*$^{-}$ and *lacY*$^{-}$ mutations are missense mutations.)

a) Haploid strain (in presence or absence of inducer). The mutant Lac repressor cannot bind to the *lacO*$^{+}$ operator, resulting in constitutive synthesis of *lac* operon enzymes.



b) Partial diploid in the absence of inducer. The *lacI*$^{+}$ operon produces wild-type Lac repressors, whereas the *lacI*$^{-}$ operon produces inactive Lac repressors. The mutant Lac repressor cannot bind to the *lacO*$^{+}$ operators, but the wild-type Lac repressors can, so no transcription occurs from either operon.

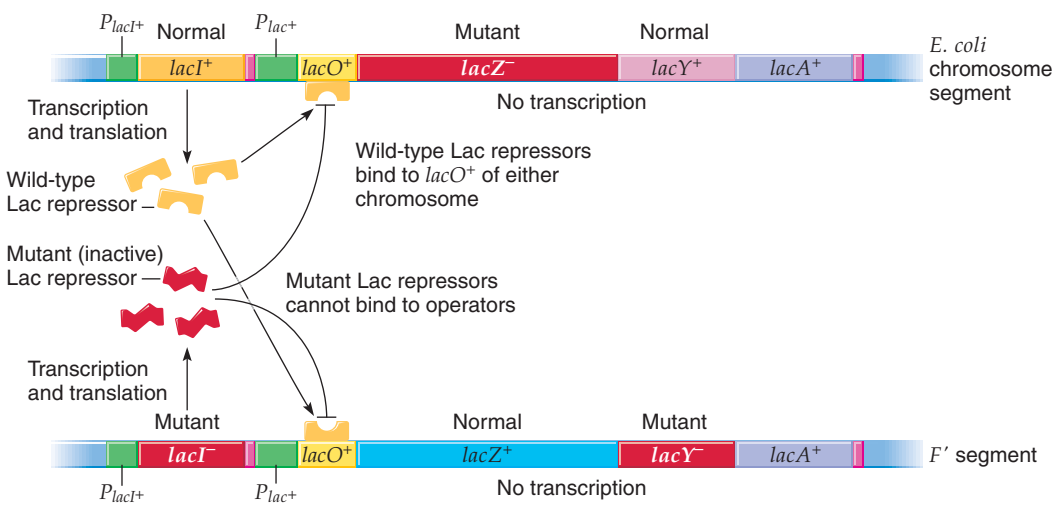


Figure 17.9 (Continued)

c) Partial diploid in the presence of inducer. The inducer inactivates the wild-type Lac repressor, preventing it from binding to the *lacO*⁺ operators. The mutant Lac repressor is unable to bind to those operators. The result is transcription of both operons: nonfunctional β-galactosidase and functional permease are produced from the *lacI*⁺ operon, and functional β-galactosidase and nonfunctional permease are produced from the *lacI*⁻ operon.

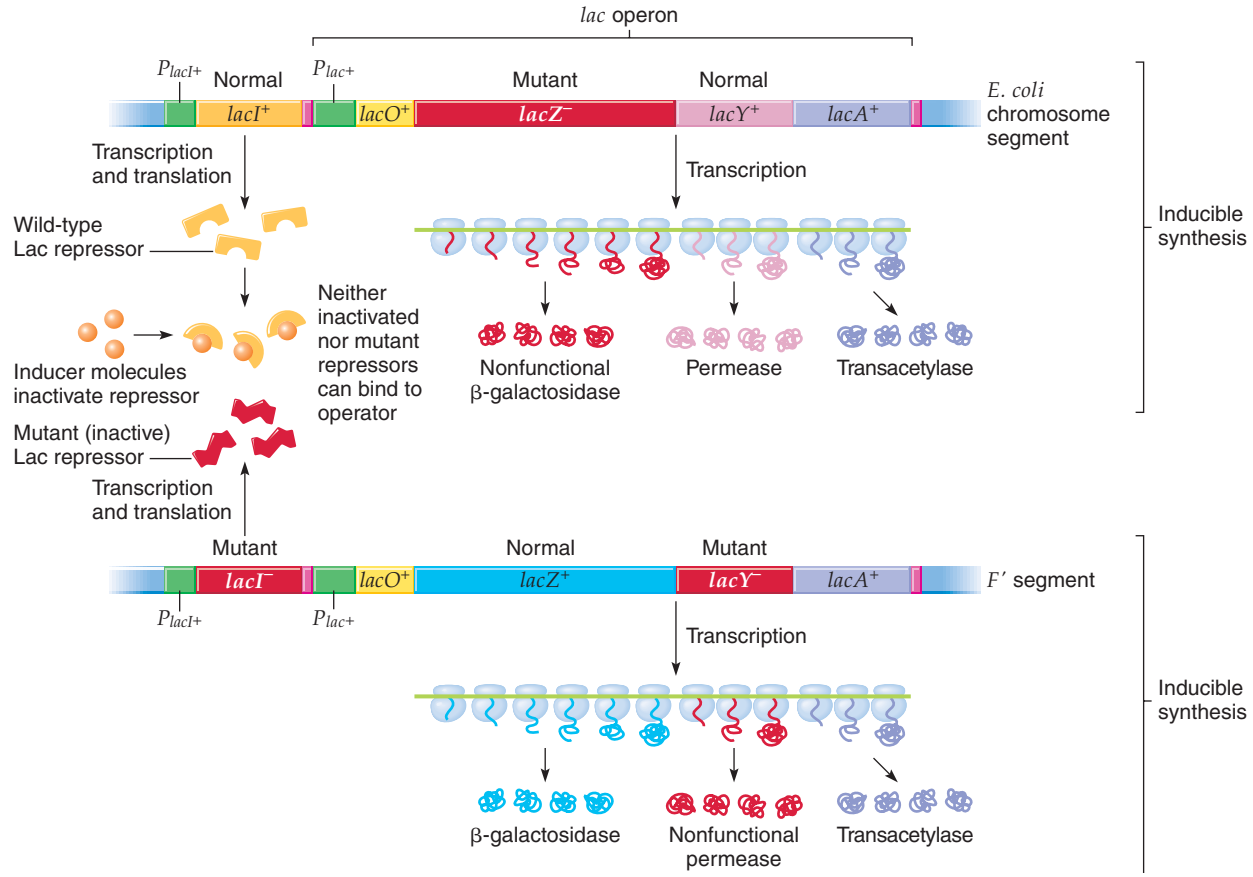
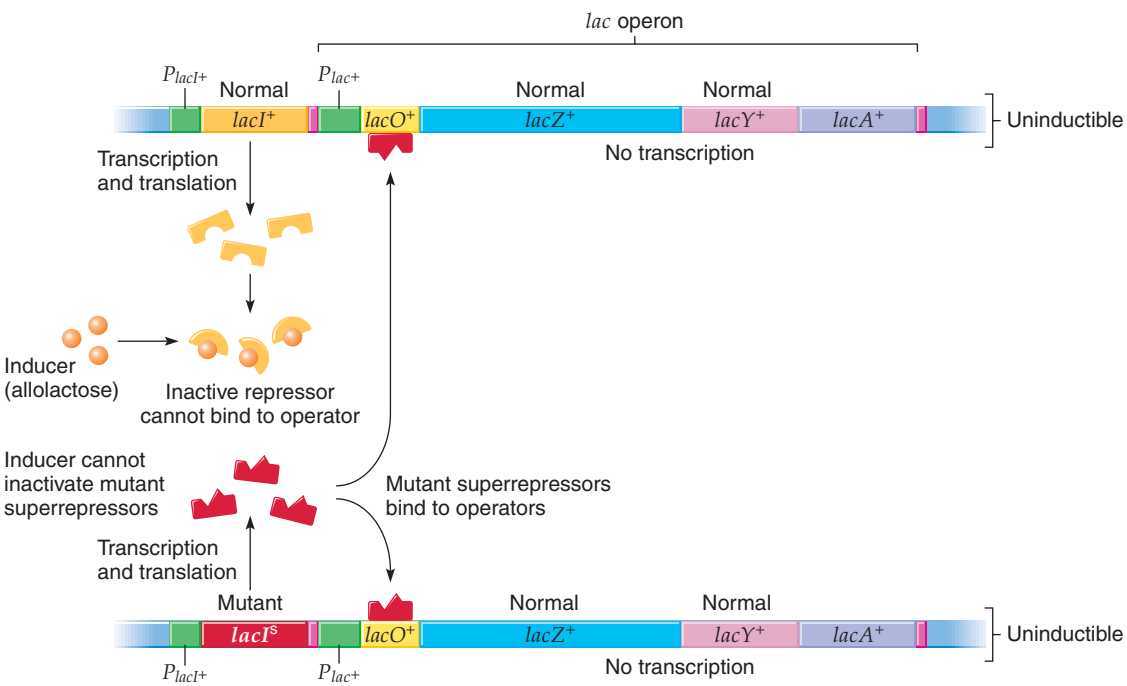


Figure 17.10

Dominant effect of *lacI*ˢ mutation over wild-type *lacI*⁺ in a *lacI*⁺ *lacO*⁺ *lacZ*⁺ *lacY*⁺ *lacA*⁺/*lacI*ˢ *lacO*⁺ *lacZ*⁺ *lacY*⁺ *lacA*⁺ partial diploid cell growing in the presence of lactose.



the operon. This system ensures that the *lac* operon will be expressed at high levels only if lactose is the sole carbon source *and not if glucose is present as well.* Glucose is a preferred carbon source because it can be used directly by the glycolytic pathway to produce "energy" for the cell. Lactose and other sugars converted to glucose consume energy. Therefore, more energy can be obtained for the cell from glucose than from other sugar sources.

Figure 17.11 shows the positive regulation of the *lac* operon if lactose is present and glucose is absent. First, a protein called **catabolite activator protein (CAP)** binds with **cAMP (cyclic AMP,** or cyclic adenosine 3′,5′-monophosphate; see Figure 17.12) to form a CAP–cAMP complex. This complex is the positive-regulator molecule. The CAP protein is a dimer of two identical polypeptides. Next, the CAP–cAMP complex binds to the *CAP site*, which is upstream of the site where RNA polymerase binds to the promoter. CAP then recruits RNA polymerase to the promoter, and transcription is initiated.

When glucose is in the medium along with lactose, the glucose is used preferentially because **catabolite repression** (also called the **glucose effect**) occurs. In catabolite repression, the *lac* operon is expressed at only very low levels even though lactose is present in the medium. This occurs because glucose causes the amount of cAMP in the cell to be reduced greatly. As a result, insufficient CAP–cAMP complex is available to recruit RNA polymerase to the *lac* promoter, and transcription is lowered significantly, even though repressors are removed from the operator by the presence of allolactose. In other words, RNA polymerase cannot bind efficiently to the promoter without the aid of the CAP–cAMP complex. That cAMP plays a crucial role in catabolite repression was shown by a number of experiments, including one in which transcription of the *lac* operon was restored by the addition of cAMP to the cell, even though glucose was still present.

The model is that catabolite repression acts on adenylate cyclase, the enzyme that makes cAMP (see Figure 17.12). In *E. coli,* adenylate cyclase is activated by the active form of an enzyme called III^{Glc}. When glucose is transported across the cell membrane into the cell, it triggers a series of events that inactivates III^{Glc}. Without active III^{Glc}, adenylate cyclase is inactivated and no new cAMP is produced. This, along with the breakdown of cAMP by phosphodiesterase, reduces the level of cAMP in the cell. Thus, cAMP is an indicator of glucose levels: when glucose levels are high, cAMP concentration is low, and when glucose levels are low, cAMP concentration is high.

Catabolite repression occurs in the same way in a number of other bacterial operons related to the catabolism of sugars other than glucose. These operons all have in common a CAP site in their promoters to which a specific CAP–cAMP complex binds to facilitate RNA polymerase binding.

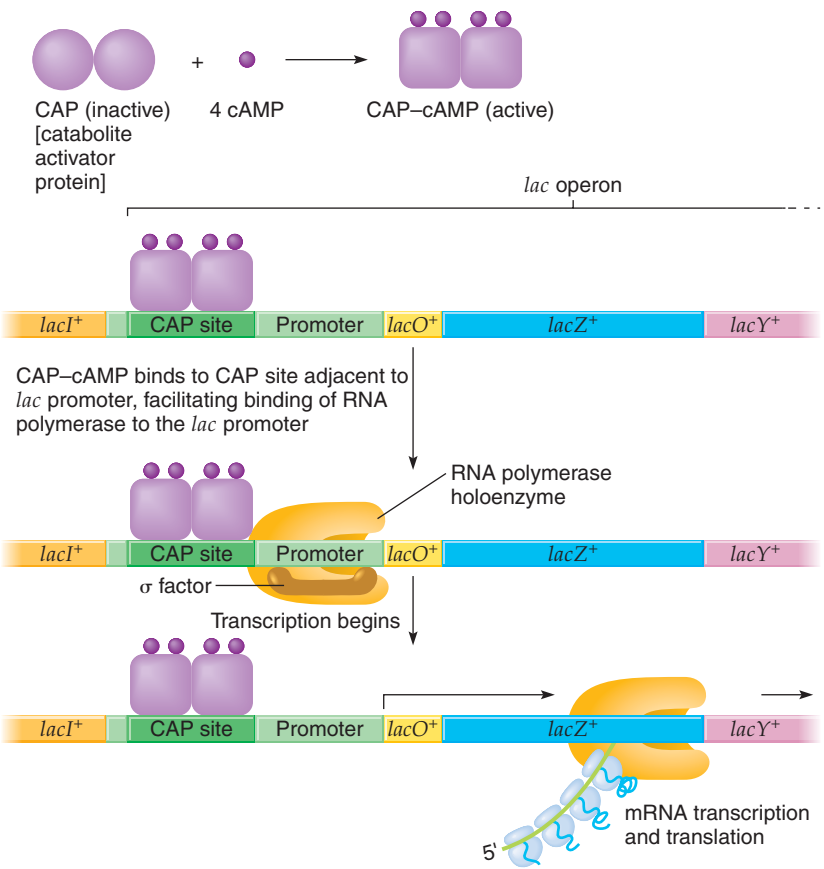
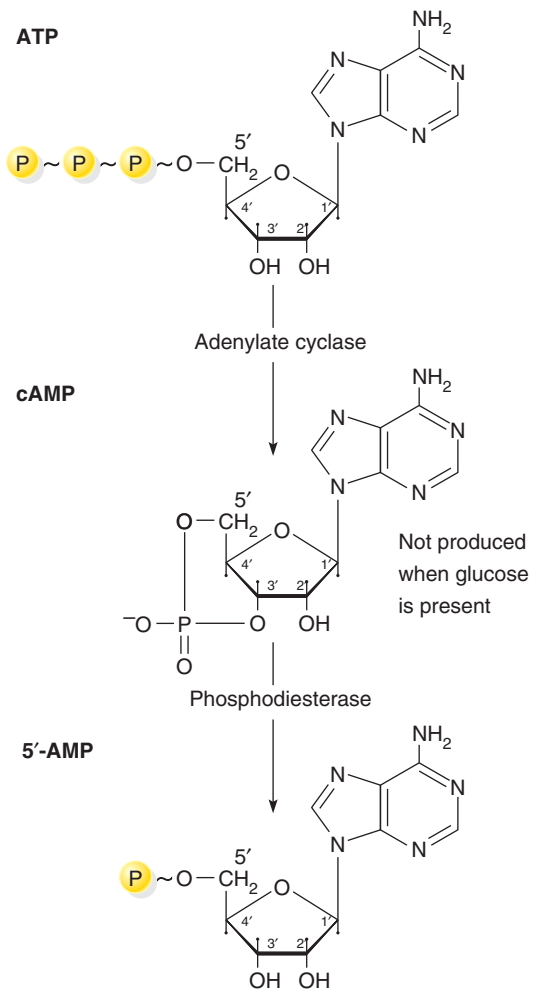


Figure 17.11

Role of cyclic AMP (cAMP) in the functioning of glucose-sensitive operons such as the *lac* operon of *E. coli*. Shown is the condition in which lactose is present and glucose is absent.

Figure 17.12

Structure, synthesis, and breakdown of cyclic AMP (cAMP, or cyclic adenosine 3′, 5′ monophosphate).



Molecular Details of *lac* Operon Regulation

From DNA- and RNA-sequencing experiments, we know the nucleotide sequences of the significant *lac* operon regulatory sequences. One general approach to obtaining this information has been to purify the protein known to bind to a regulatory sequence and to let it bind to isolated *lac* operon DNA *in vitro*. For example, if the repressor is bound to the *lac* operator, it will protect that region of the operon from deoxyribonuclease digestion. If DNase is allowed to digest the rest of the DNA, the intact operator sequence, cloned by using recombinant DNA technology, and sequenced.

Promoter Region of the *lac* Repressor Gene (*lacI*). Figure 17.13 shows the nucleotide pair sequence for the *lacI* gene promoter region, the sequence for the 5′ end of the repressor mRNA, and the first few amino acids of the repressor protein itself. The nucleotide sequence of the repressor mRNA can be aligned with this promoter sequence, with its start approximately in the middle. As with all gene transcripts, translation does not start right at the end of the mRNA molecule. The ribosome binding site is a Shine–Dalgarno sequence (AGGG) at 12 to 9 bases upstream from the start codon (Chapter 6, p. 115). In this unusual case, the start codon is GUG rather than AUG, at nucleotides 27 to 29 from the 5′ end of the messenger. The figure also shows the single base-pair change found for a particular $lacI^{Q}$ mutant; this change, from C-G to T-A, brings about a tenfold increase in repressor production.

***lac* Operon Regulatory Sequences.** Figure 17.14 shows the nucleotide pair sequence of the *lac* operon regulatory sequences. The orientation of this sequence was put together from several different pieces of information. First, the amino acid sequences of the repressor protein and of β-galactosidase were completely known, and that information made it possible to identify the coding regions of the *lacI* gene and of the $lacZ^{+}$ gene. Then the other regions were identified on the basis of "protection" experiments of the kind described previously. Here, CAP–cAMP complex, RNA polymerase, and repressor protein were used separately to bind to the DNA, and DNase-resistant regions were then sequenced.

The beginning of the promoter region is defined as position −84 in the figure (i.e., 84 base pairs upstream from the mRNA initiation site), immediately next to the stop codon for the *lacI* gene. The consensus sequence matches for the CAP–cAMP binding site are nucleotide pairs −54 to −58 and −65 to −69, and the DNA covered

Figure 17.13

Base-pair sequences of the *lac* operon $lacI^{+}$ gene promoter (P_{lac}^{+}) and of the 5′ end of the repressor mRNA. Also shown is the amino acid sequence of the first part of the repressor protein itself. Note that GUG is the initiation codon for methionine in this case.

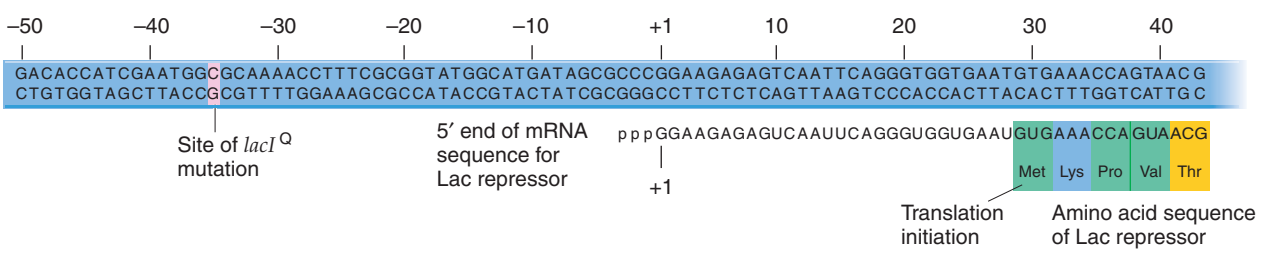
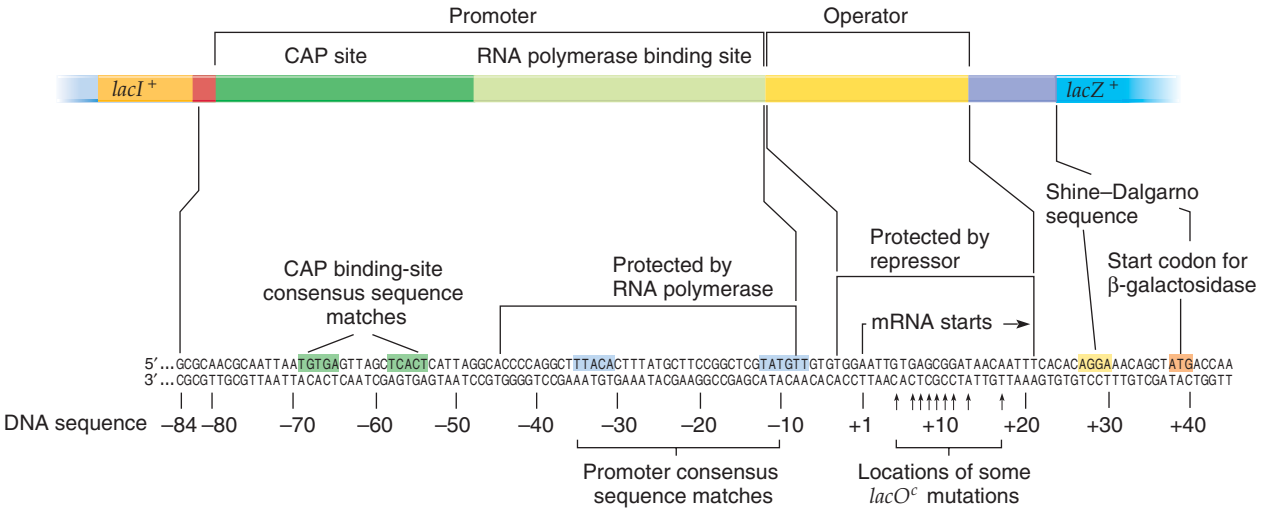


Figure 17.14

Base-pair sequence of the promoter and operator for the *lac* operon of *E. coli*.



by RNA polymerase spans nucleotide pairs −44 to −8, including −10 and −35 consensus sequence matches (see Figure 17.11). Together, the region from −84 to −8, which includes the CAP protein and the RNA polymerase interaction sites (including a Pribnow box), essentially defines the *lac* operon promoter region.

Adjacent to the promoter region is the operator. The region protected by the Lac repressor protein is the area containing nucleotide pairs −3 to +21. When the Lac repressor is bound to the operator, RNA polymerase cannot bind to the promoter.

The β-galactosidase mRNA has a *leader region* before the start codon is encountered. The actual start of the mRNA here is nucleotide pair +1 in Figure 17.14, which is very close to the beginning of the repressor binding site. Transcription of the *lac* operon includes a large proportion of the operator region, in addition to the protein-coding genes themselves. The AUG start codon for β-galactosidase, which defines the beginning of the *lacZ* coding sequence, is at nucleotide pairs +39 to +41. Thus, the first 38 bases of the *lac* mRNA are not translated.

Figure 17.14 also shows the sites of base-pair substitutions that have been identified for some of the $lacO^c$ mutations studied. In each case, a single base-pair change is responsible for the altered regulatory control of the *lac* operon.

In conclusion, the *lac* operon has proved to be a model system for understanding gene regulation in prokaryotic organisms. Jacob and Monod's original work on this system had a great impact on further studies. As the first molecular model for the regulation of gene expression in any organism, it sparked numerous studies in both prokaryotes and eukaryotes to see whether operons were generally present. We now know that operons are prevalent in bacteria and bacteriophages, but they are very rarely encountered in eukaryotes.

Activity

Your job is to determine the location and effect of a mutation in strains of *E. coli* in the iActivity *Mutations and Lactose Metabolism* on the student website.

Keynote

Studies of the synthesis of the lactose-utilizing enzymes of *E. coli* generated a model that is the basis for the regulation of gene expression in a large number of bacterial and bacteriophage systems. In the lactose system, the addition of lactose to the cells brings about a rapid synthesis of three enzymes. The genes for these enzymes are contiguous on the *E. coli* chromosome and are adjacent to two regulatory sequences: a promoter and an operator. The promoter, operator, and genes constitute an operon, which is transcribed as a single unit. In the absence of lactose, the operon is turned off by a repressor.

A positive control system also regulates the *lac* operon. That is, CAP–cAMP binds to the promoter, and this facilitates the binding of RNA polymerase to the promoter. If glucose is present, however, no CAP–cAMP is produced, so RNA polymerase cannot bind efficiently and the *lac* genes are not transcribed.

The *trp* Operon of *E. coli*

E. coli has certain operons and other gene systems that enable it to manufacture any amino acid that is lacking in the medium in which it is placed, so that it can grow and reproduce. When an amino acid is present in the growth medium, though, the genes encoding the enzymes for biosynthetic pathway for that amino acid are turned off. Unlike the *lac* operon, wherein gene activity is induced when a chemical (lactose) is added to the medium, in this case gene activity is repressed when a

chemical (an amino acid) is added. We call amino acid biosynthesis operons controlled in this way **repressible operons.** In general, operons for anabolic (biosynthetic) pathways are repressed (turned off) when the end product is readily available. One repressible operon in *E. coli* that has been extensively studied is the operon for the biosynthesis of the amino acid tryptophan (Trp).

Gene Organization of the Tryptophan Biosynthesis Genes

Figure 17.15 shows the organization of the regulatory sequences and of the genes that code for the tryptophan biosynthetic enzymes and how they relate to the biosynthetic steps. Much of the work we will discuss is that of Charles Yanofsky and his collaborators.

Five structural genes (*A–E*) occur in the *trp* operon. The promoter and operator regions are upstream from the *trpE* gene. Between the promoter–operator region and *trpE* is a short region called *trpL*, the leader region. Within *trpL*, close to *trpE*, is an *attenuator site* (*att*) that plays an important role in the regulation of the *trp* operon.

The entire *trp* operon is approximately 7,000 base pairs long. Transcription of the operon results in the production of a polycistronic mRNA for the five structural genes.

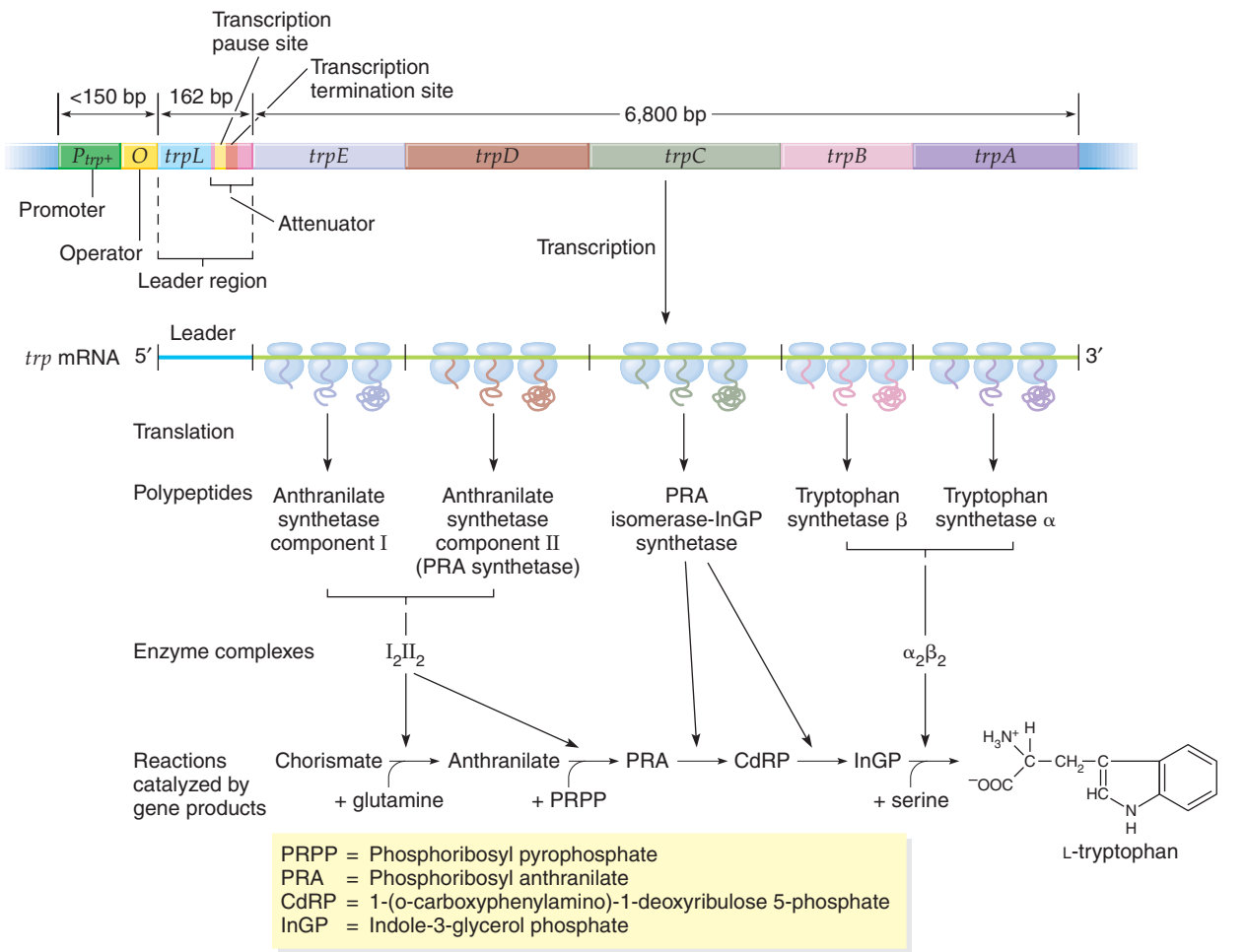
Regulation of the *trp* Operon

Two regulatory mechanisms are involved in controlling the expression of the *trp* operon. One mechanism uses a repressor–operator interaction, and the other determines whether initiated transcripts include the structural genes or are terminated before those genes are reached.

Expression of the *trp* Operon in the Presence of Tryptophan. The regulatory gene for the *trp* operon is *trpR*, located some distance from the operon (and therefore not shown in Figure 17.15). The product of *trpR* is an **aporepressor protein,** which is basically an inactive repressor that alone cannot bind to the operator. When tryptophan is abundant within the cell, it interacts with the aporepressor and converts it to an active Trp repressor. (Tryptophan is an example of an effector molecule,

Figure 17.15

The regulatory sequences and structural genes of the *E. coli trp* operon, and the functions of the structural gene products.



just as allolactose is the effector molecule for the *lac* repressor.) The active Trp repressor binds to the operator and prevents the initiation of transcription of the *trp* operon protein-coding genes by RNA polymerase. As a result, the tryptophan biosynthesis enzymes are not produced. By repression, transcription of the *trp* operon can be reduced about seventy-fold.

Expression of the *trp* Operon in the Presence of Low Concentrations of Tryptophan. The second regulatory mechanism is involved in the expression of the *trp* operon under conditions of tryptophan starvation or tryptophan limitation. Under severe tryptophan starvation, the *trp* genes are expressed maximally; under less severe starvation conditions, the *trp* genes are expressed at less than maximal levels. This is accomplished by a mechanism that controls the ratio of full-length transcripts that include the five *trp* structural genes to short, 140-bp transcripts that have terminated at the attenuator site within the *trpL* region (see Figure 17.15). The short transcripts are terminated by a process called **attenuation.** The proportion of the transcripts that include the structural genes is inversely related to the amount of tryptophan in the cell; the more tryptophan there is, the greater is the proportion of short transcripts. Attenuation can reduce transcription of the *trp* operon by a factor of 8 to 10. Thus, repression and attenuation together can regulate the transcription of the *trp* operon by a factor of about 560 to 700.

Molecular Model for Attenuation. The mRNA transcript of the leader region includes a sequence that can be translated to produce a short polypeptide. Just before the stop codon in the transcript are two adjacent codons for tryptophan that play an important role in attenuation.

There are four regions of the leader peptide mRNA that can fold and form secondary structures by complementary base pairing (Figure 17.16). The pairing of regions 1 and 2 results in a transcription *pause signal*, that of 3 and 4 is a *termination of transcription signal*, and the pairing of 2 and 3 is an *antitermination signal* for transcription to continue.

animation
Attenuation in the *trp* Operon of *E. coli*

Crucial to the attenuation model is the fact that transcription and translation are tightly coupled in prokaryotes, made possible by the absence of a nuclear envelope and the lack of processing of mRNA transcripts. In the *trp* regulatory system, a pause of the RNA polymerase is caused by the pairing of RNA regions 1 and 2 just after they have been synthesized (see Figure 17.16). The pause enables the ribosome to load onto the mRNA and to begin translating the leader peptide so that translation of the leader mRNA transcript occurs just behind transcription by RNA polymerase.

As coupled transcription and translation continues, the position of the ribosome on the leader transcript plays an important role in the regulation of transcription termination at the attenuator. If the cells are starved for tryptophan, the amount of Trp–tRNA molecules (charged tryptophanyl–tRNA) drops dramatically, since very few tryptophan molecules are available for the aminoacylation of the tRNA. A ribosome translating the leader transcript stalls at the tandem Trp codons in region 1 because the next specified amino acid in the peptide is in short supply; the leader peptide cannot be completed (Figure 17.17a). Since the ribosome now "covers" region 1 of the attenuator region, the 1:2 pairing cannot happen, as region 1 is no longer available. However, RNA region 2 will pair with RNA region 3

Figure 17.16
Four regions of the *trp* operon leader mRNA and the alternative secondary structures they can form by complementary base pairing.

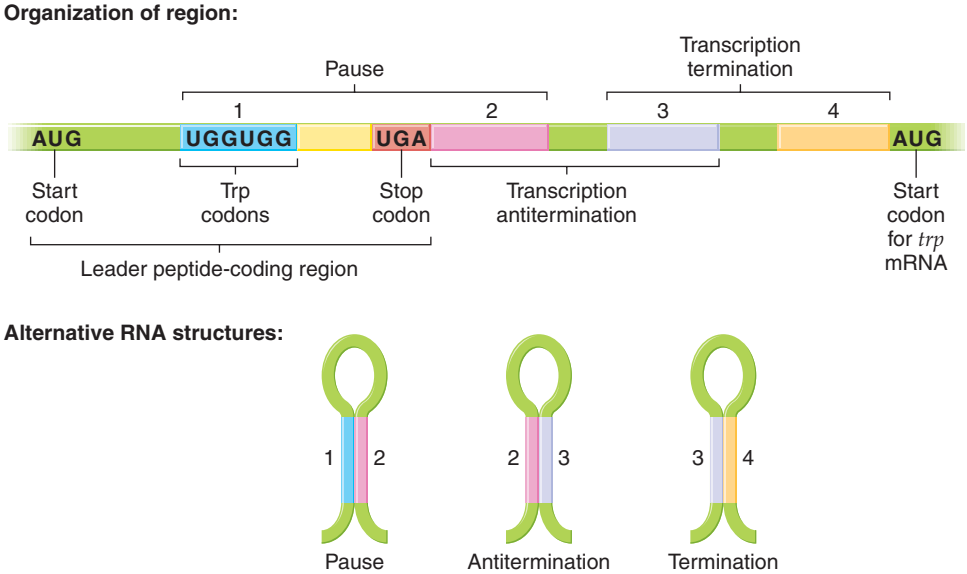
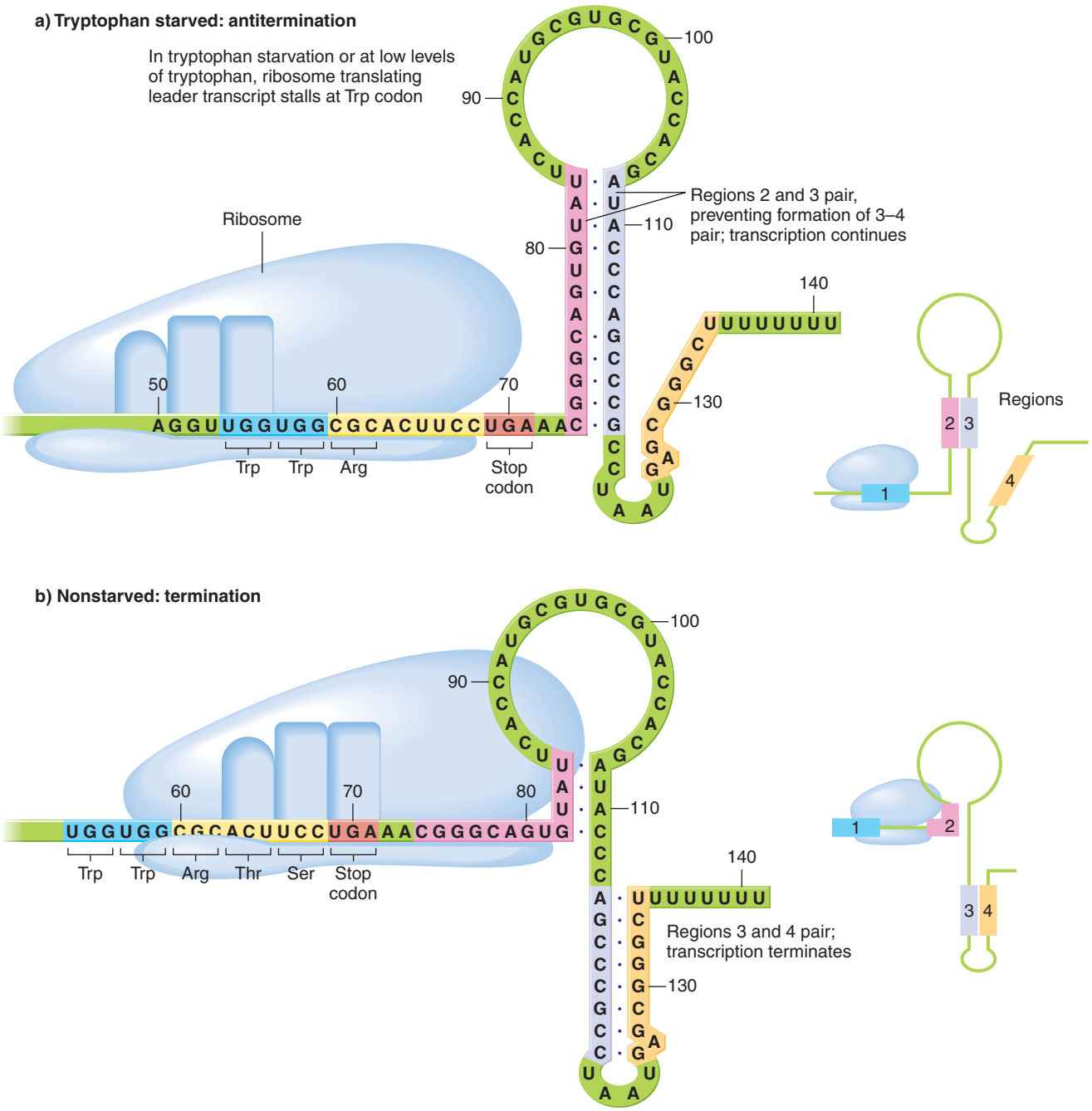


Figure 17.17

Model for attenuation in the *trp* operon of *E. coli*.


a) Tryptophan starved: antitermination
In tryptophan starvation or at low levels of tryptophan, ribosome translating leader transcript stalls at Trp codon
Ribosome
Regions 2 and 3 pair, preventing formation of 3–4 pair; transcription continues
AGGUUGGUGGCGCACUUCCUGAAAC
Trp Trp Arg
Stop codon
Regions
b) Nonstarved: termination
UGGUGGCGCACUUCCUGAAACGGGCAGUG
Trp Trp Arg Thr Ser Stop codon
Regions 3 and 4 pair; transcription terminates

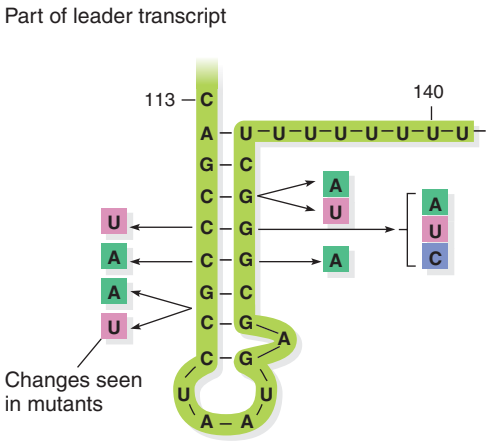
once region 3 is synthesized. Because region 3 is paired with region 2, region 3 cannot pair with region 4 when it is synthesized. The 2:3 pairing is an antitermination signal, since the termination signal of 3 paired with 4 does not form, thereby allowing RNA polymerase to continue past the attenuator and transcribe the structural genes.

If, instead, enough tryptophan is present so that the ribosome can translate the Trp codons (Figure 17.17b), then the ribosome continues to the stop codon for the leader peptide. Since the ribosome is then covering part of RNA region 2, that region is unable to pair with region 3, and region 3 is then able to pair with region 4 when it is transcribed. The bonding of region 3 with region 4 is a transcription termination signal (see rho-independent terminator, Figure 5.5, p. 86). The 3:4 structure is called the *attenuator.* The key signal for attenuation is the concentration of Trp–tRNA in the cell because that determines how far the ribosome gets on the leader transcript, either to the Trp codons or to the stop codon.

Genetic evidence for the attenuation model has been obtained through the study of mutants. One type of mutant shows less efficient transcription termination at the attenuator, increasing structural gene expression. The mutations involved are single base-pair changes leading

Figure 17.18

In the *trpL* region, mutation sites that show less efficient transcription at the attenuator site. The mutations map to DNA regions that correspond to regions 3 and 4 in the RNA.



to the base changes in the leader transcript shown in Figure 17.18. In each case, the change is in the regions of 3:4 pairing; each causes a disruption of the pairing so that the structure is less stable. In the less stable state, the structure is less able to prevent transcription from proceeding into the structural genes.

Further direct evidence for the attenuation model came from DNA manipulations in which the DNA sequences for the two Trp codons were changed to encode another amino acid. In those mutant strains, attenuation was not seen in response to changing levels of tryptophan, but it was seen in response to changing levels of the amino acid now specified by the codons.

Attenuation is involved in the genetic regulation of a number of other amino acid biosynthetic operons of *E. coli* and *Salmonella typhimurium*. In every case, there is a leader sequence with two or more codons for the particular amino acid, the synthesis of which is controlled by the enzymes encoded by the operon (Figure 17.19). For example, the *his* operon of *E. coli* has a string of seven histidines in the leader peptide, and 7 of the 15 amino acids in the leader for the *pheA* operon of *E. coli* are phenylalanine. Attenuation has also been shown to regulate a number of genes not involved with amino acid biosynthesis, such as the *ampC* gene of *E. coli* (for resistance to ampicillin).

Keynote

Regulation of the tryptophan (*trp*) operon of *E. coli* is at the level of initiating completing a transcript of the operon. This is accomplished through a repressor–operator system, which responds to free tryptophan levels, and through attenuation at a second regulatory sequence called an attenuator, which responds to Trp–tRNA levels. The attenuator is located in the leader region between the operator region and the first *trp* structural gene. The attenuator acts to terminate transcription, depending on the concentration of tryptophan. In the presence of large amounts of tryptophan, attenuation is highly effective; that is, enough Trp–tRNA is present so that the ribosome can move past the attenuator and allow the leader transcript to form a secondary structure that causes transcription to be blocked. In the absence of tryptophan or at low amounts of the amino acid, the ribosomes stall at the attenuator and the leader transcript forms a secondary structure that permits transcription to continue.

The *ara* Operon of *E. coli*: Positive and Negative Control

Jacob and Monod's operon model for the regulation of gene expression convinced most researchers of the time that negative control was involved in the regulation of gene expression in all systems. In the *lac* and *trp* operons discussed earlier, the negative control is the result of the action of a repressor blocking transcription. In the case of the *lac* operon, the repressor produced by translation of the mRNA transcribed from *lacI* is active, blocking transcription in that form. When the inducer is present, the repressor is inactivated, allowing transcription of the structural genes to

Figure 17.19

Predicted amino acid sequences of the leader peptides of a number of attenuator-controlled bacterial operons. Shown are leader peptides for the *pheA, his, leu, thr,* and *ilv* (isoleucine and valine) operons of *E. coli* or *Salmonella typhimurium*. The amino acids that regulate the respective operons are highlighted in orange.

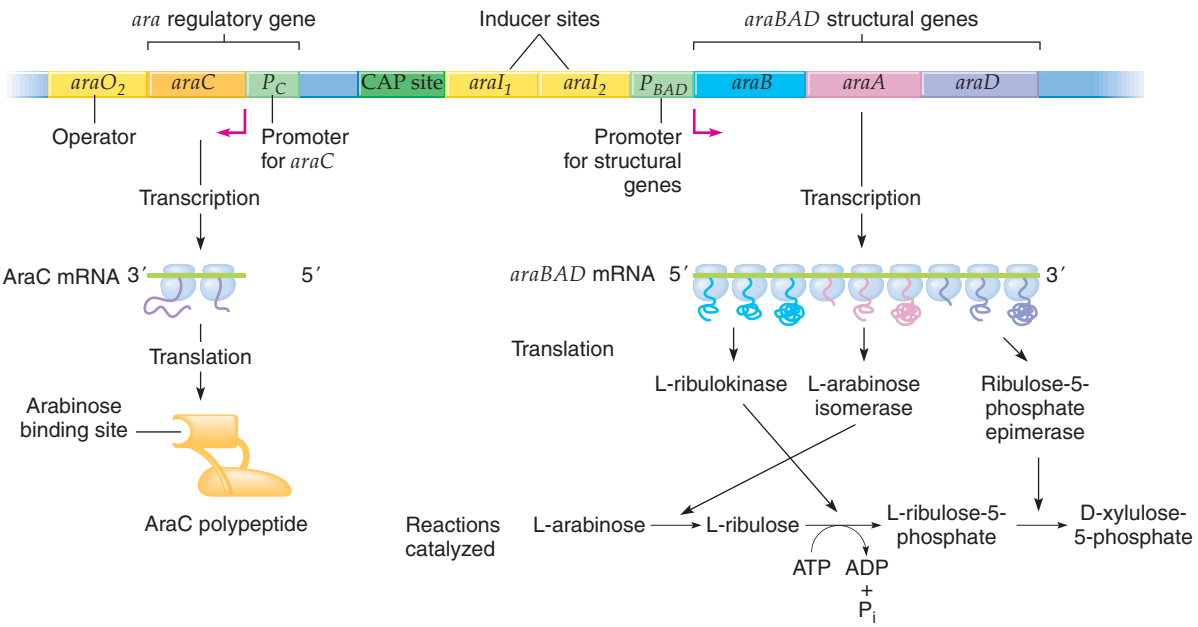
Operon	Leader peptide sequence
pheA:	Met – Lys – His – Ile – Pro – Phe – Phe – Phe – Ala – Phe – Phe – Phe – Thr – Phe – Pro – –
his:	Met – Thr – Arg – Val – Gln – Phe – Lys – His – His – His – His – His – His – His – Pro – Asp – –
leu:	Met – Ser – His – Ile – Val – Arg – Phe – Thr – Gly – Leu – Leu – Leu – Leu – Asn – Ala – Phe – Ile – Val – Arg – Gly – Arg – Pro – Val – Gly – Ile – Gln – His – –
thr:	Met – Lys – Arg – Ile – Ser – Thr – Thr – Ile – Thr – Thr – Thr – Ile – Thr – Ile – Thr – Thr – Gly – Asn – Gly – Ala – Gly – –
ilv:	Met – Thr – Ala – Leu – Leu – Arg – Val – Ile – Ser – Leu – Val – Val – Ile – Ser – Val – Val – Val – Ile – Ile – Ile – Pro – Pro – Cys – Gly – Ala – Ala – Leu – Gly – Arg – Gly – Lys – Ala – –

occur. In the case of the *trp* operon, the repressor is inactive on its own, becoming active in blocking transcription only when bound to tryptophan. Also in the 1960s, at about the same time Jacob and Monod were doing their experiments, Ellis Englesberg and his coworkers were using genetic analysis to study the regulation of the arabinose (*ara*) operon of *E. coli*. Their results indicated that the *ara* operon was under positive control; that is, an *activator* is needed for transcription to occur. While that conclusion was not immediately accepted at the time, subsequent biochemical and molecular analysis supported it. Positive regulation involving activators is now known to occur in a variety of prokaryotic systems (including the *lac* operon, as already detailed) and in all eukaryotes.

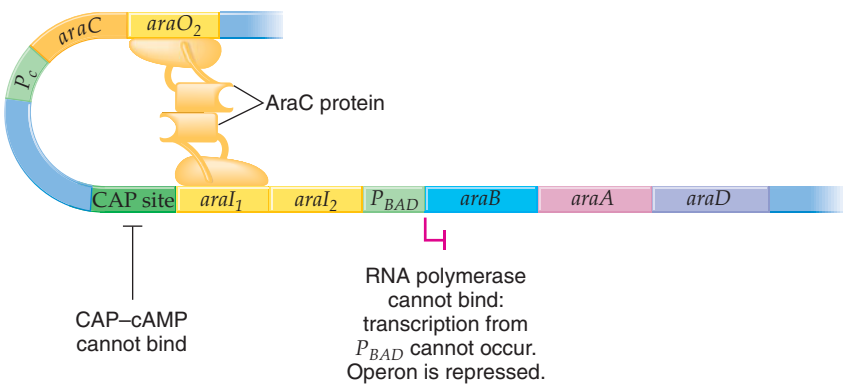
Figure 17.20a shows the organization of the *ara* operon. The three structural genes, *araB*, *araA*, and *araD*,

Figure 17.20
Regulation of the *ara* operon of *E. coli*.

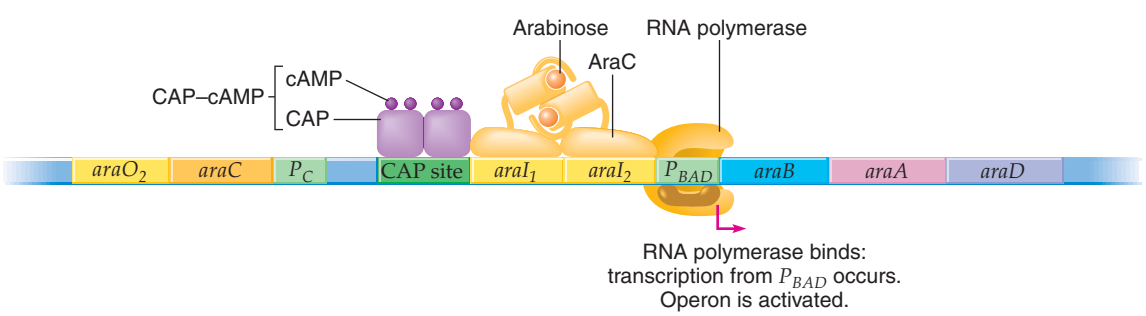
a) Organization of the *ara* operon



b) Arabinose and glucose absent: transcription of structural genes is inhibited



c) Arabinose present and glucose absent: transcription of structural genes is induced



encode three enzymes that convert the pentose sugar L-arabinose to D-xylulose-5-phosphate, which is then metabolized by other biochemical pathways. Like lactose, arabinose can be used as the sole source of carbon and energy for *E. coli*. In this case, arabinose is the inducer of the operon. And, also like the *lac* operon and other sugar utilization operons, this operon can be transcribed only if glucose is absent, because cAMP–CAP must bind to the CAP site in order for RNA polymerase to bind (see "Positive Control of the *lac* Operon," pp. 499–501). Upstream of the structural genes is the regulatory gene, *araC*, and several controlling sites for transcription of that gene and the structural genes. The polypeptide product of the regulatory gene forms a dimer of two identical subunits, named AraC; it is the key regulator of gene expression. P_{BAD} is the promoter where RNA polymerase binds to transcribe the structural genes (from left to right), and P_C is the promoter where RNA polymerase binds to transcribe *araC* (from right to left). The other sites will be explained as the regulatory events are described.

When arabinose and glucose are absent from the growth medium, the *araBAD* structural genes are not transcribed (Figure 17.20b). This is accomplished by one subunit of the AraC protein binding to the inducer site, $araI_1$, and the other subunit binding to the operator, $araO_2$. Binding of AraC in this way causes the DNA to form a loop. The loop blocks CAP–cAMP from binding to the CAP site and RNA polymerase from binding to P_{BAD}; transcription of the structural genes therefore does not occur. This part of the regulation of the *ara* operon involves negative control in which AraC acts as a repressor.

When arabinose is present in the growth medium (and glucose is absent), the *ara* operon is induced (Figure 17.20c). The switch is thrown to transcribing the *araBAD* genes when arabinose, acting as an inducer, binds to each subunit of AraC, causing an allosteric shift in the protein. In the changed conformation, one subunit of AraC remains bound to the $araI_1$ site, but the other subunit releases from $araO_2$ and binds instead to $araI_2$. As a result, the DNA no longer forms a loop. Therefore, CAP–cAMP binds to the CAP site, and RNA polymerase binds to P_{BAD} and transcribes the structural genes. This part of the regulation of the *ara* operon involves positive control in which AraC acts as an activator.

If glucose is present as well as arabinose, the *araBAD* genes are not transcribed due to catabolite repression. That is, as for the *lac* operon, the decrease in cAMP levels that occurs when glucose is present means CAP–cAMP complexes are not formed. In the absence of CAP–cAMP binding to the CAP site, initiation of transcription by RNA polymerase at P_{BAD} is highly inefficient.

Regulation of Gene Expression in Phage Lambda

Bacteriophages exist by invading and manipulating bacterial cells. Many or all of the essential components for phage reproduction are provided by the bacterial host cell, and the use of those components is controlled by the products of phage genes. Most genes of a phage, then, code for products that control the life cycle and the production of progeny phage particles. Much is known about gene regulation in a number of bacteriophages. In this section, we discuss the regulation of gene expression as it relates to the lytic cycle and lysogeny in bacteriophage lambda (λ). (Recall from Chapter 15, p. 440–442, that in the lytic cycle, the phage takes over the bacterium and directs its growth and reproductive abilities so that it expresses the phage's genes and produces progeny phages. Lysogeny involves the insertion of a temperate phage chromosome into a bacterial chromosome, with the former replicating whenever the latter does. In this state, the phage genome is repressed and is said to be in the *prophage* state.)

Early Transcription Events

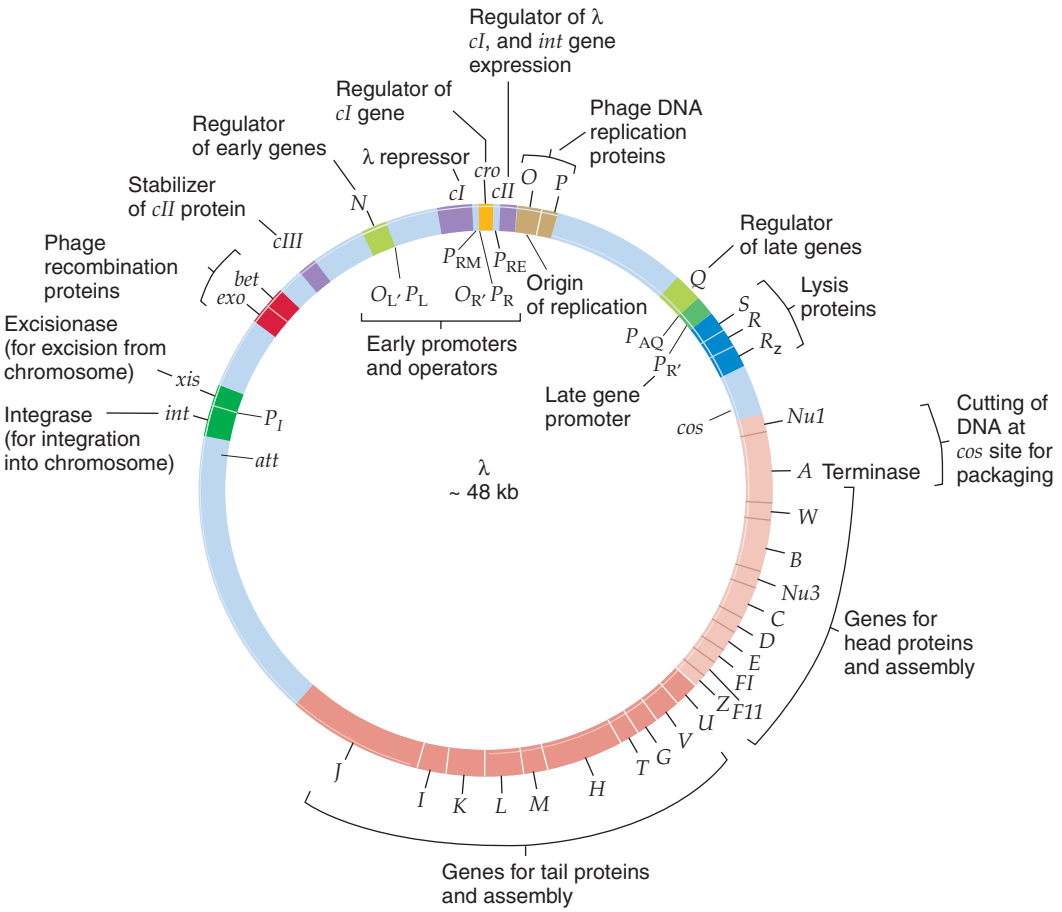
Figure 17.21 shows the genetic map of λ. The mature λ chromosome is linear and has complementary "sticky" ends. Once free in the host cell, the λ chromosome circularizes, so we show the genetic map in a circular form. Recall that λ is a temperate phage (see Figure 15.12, p. 441), so when it infects a bacterial cell, the phage has a choice of whether to enter the lytic pathway (when progeny phages are assembled and released from the cell) or the lysogenic pathway (when the λ chromosome integrates into the chromosome and no progeny phages are produced). The regulatory system involved in this choice is an excellent model for a genetic switch and, as such, has contributed to our thinking about how genetic switches might operate in eukaryotic systems.

The choice between the lytic and lysogenic pathways occurs soon after λ infects the cell and its genome circularizes. The choice involves a sophisticated *genetic switch*. First, transcription begins at promoters P_L and P_R (Figure 17.22, part 1). Promoter P_L is for leftward transcription of the left early operon, and promoter P_R is for rightward transcription of the right early operon.

The first gene to be transcribed from P_R is *cro* (control of repressor and other), the product of which is the Cro protein. This protein plays an important role in setting the genetic switch to the lytic pathway. The first gene to be transcribed from P_L is *N*. The resulting N protein is a transcription *antiterminator* that allows RNA synthesis to proceed past certain transcription terminators, in this case leftward of *N* and rightward of *cro*, thereby including all the early genes (Figure 17.22, part 2). Genes transcribed due to the action of the N protein are *cII*, *O*, *P*, and Q. Gene *cII* encodes protein cII, which can turn on gene *cI* (which encodes the λ repressor) and gene *int* (which encodes the integrase required for integrating the lambda chromosome into the host chromosome during the lysogenic pathway). However, cII protein performs this function only when the phage follows the lysogenic pathway. Genes *O* and *P* encode two DNA replication proteins, and gene *Q* encodes a protein needed to turn on late genes for

Figure 17.21

A map of phage λ, showing the major genes. (Promoters discussed in text: P_L = promoter for leftward transcription of the left early operon, P_R = promoter for rightward transcription of the right early operon, P_{RE} = promoter for repressor establishment, and P_{RM} = promoter for repressor maintenance.)



lysis and phage particle proteins. The Q protein is another antiterminator, permitting transcription to continue into the late genes involved in the lytic pathway. However, only when the switch is set to the lytic pathway and transcription continues from P_R for a sufficient time does enough Q protein accumulate to function effectively.

The Lysogenic Pathway

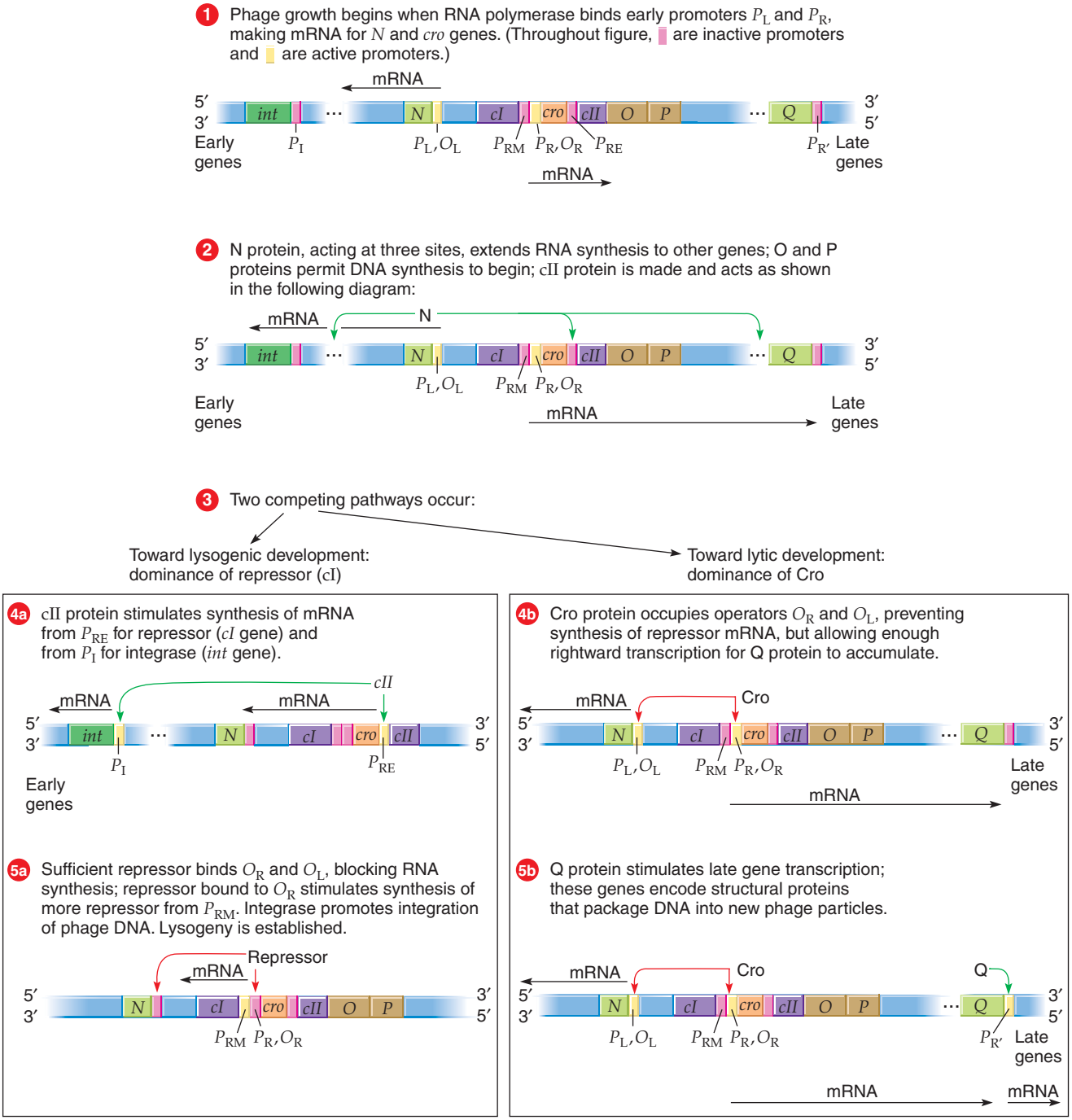
After the early transcription events, either the lysogenic or lytic pathway is followed (Figure 17.22, part 3). The switch is set for the lysogenic pathway as follows.

The establishment of lysogeny requires the protein products of the *cII* (right early operon) and *cIII* (left early operon; see Figure 17.21) genes. The cII protein (stabilized by cIII protein) activates transcription of the *cI* gene (located between the P_L and P_R promoters; see Figure 17.22, part 4a) leftward from a promoter called P_{RE} (promoter for repressor establishment). The *cro* gene is not transcribed during this event, however, because transcription of *cro* occurs in the rightward direction under the control of a promoter on its left. The product of the *cI* gene, the λ repressor, binds to two operator regions, O_L and O_R (see Figure 17.21 and Figure 17.22, part 5a), whose sequences overlap the P_L and P_R promoters, respectively. The binding of the λ repressor prevents the further transcription by RNA polymerase of the early operons controlled by P_L and P_R. As a result, transcription of the *N* and *cro* genes is blocked, and because the two proteins specified by these genes are unstable, the concentrations of those two proteins in the cell drop dramatically. Furthermore, a repressor bound to O_R stimulates the synthesis of more repressor mRNA from a different promoter, P_{RM} (promoter for repressor maintenance), thereby maintaining repressor concentrations in the cell (see Figure 17.22, part 5a). Thus, if enough λ repressors are present, lysogeny is established by the binding of the repressor to operators O_L and O_R, followed by the integration of λ DNA catalyzed by integrase, which is the product of the cII-regulated promoter P_I. As the concentration of cII drops, P_I transcription shuts off, leaving P_{RM} as the only active promoter.

In sum, the lysogenic pathway is favored when enough λ repressor is made so that early promoters are turned off, thereby repressing all the genes needed for the

Figure 17.22

Expression of λ genes after infection of *E. coli*, and the transcriptional events that occur when either the lysogenic or lytic pathway is followed. In the figure, stimulation of transcription is indicated by green arrows, and repression of transcription by red arrows.


1 Phage growth begins when RNA polymerase binds early promoters P_L and P_R, making mRNA for N and *cro* genes. (Throughout figure, ▮ are inactive promoters and ▮ are active promoters.)
mRNA
5′ 3′
int N cI cro cII O P Q
3′ 5′
Early genes
P_I P_L,O_L P_{RM} P_R,O_R P_{RE} $P_{R'}$
Late genes
mRNA
2 N protein, acting at three sites, extends RNA synthesis to other genes; O and P proteins permit DNA synthesis to begin; cII protein is made and acts as shown in the following diagram:
mRNA
N
Early genes
Late genes
mRNA
3 Two competing pathways occur:
Toward lysogenic development: dominance of repressor (cI)
Toward lytic development: dominance of Cro
4a cII protein stimulates synthesis of mRNA from P_{RE} for repressor (*cI* gene) and from P_I for integrase (*int* gene).
cII
mRNA
mRNA
Early genes
4b Cro protein occupies operators O_R and O_L, preventing synthesis of repressor mRNA, but allowing enough rightward transcription for Q protein to accumulate.
mRNA
Cro
Late genes
mRNA
5a Sufficient repressor binds O_R and O_L, blocking RNA synthesis; repressor bound to O_R stimulates synthesis of more repressor from P_{RM}. Integrase promotes integration of phage DNA. Lysogeny is established.
Repressor
mRNA
5b Q protein stimulates late gene transcription; these genes encode structural proteins that package DNA into new phage particles.
mRNA
Cro
Q
Late genes
mRNA
mRNA

lytic pathway. One important lytic pathway gene that is repressed is *Q;* the Q protein is a positive regulatory protein required for the production of phage coat proteins and lysis proteins (see the next section).

The Lytic Pathway

Let us consider the induction of the lytic pathway caused by ultraviolet light irradiation. Inducers such as ultraviolet light typically damage DNA, and this somehow causes a change in the function of the bacterial protein RecA (the product of the *recA* gene). Normally RecA functions in DNA recombination, but when DNA is damaged, RecA stimulates the λ repressor polypeptides to cleave themselves in two and therefore become inactivated. The resulting absence of repressor at O_R allows RNA polymerase to bind at P_R, and the *cro* gene is then further transcribed. The Cro protein that is produced then acts to decrease

RNA synthesis from P_L and P_R, and this reduces the synthesis of the cII protein, the regulator of λ repressor synthesis, and blocks the synthesis of λ repressor mRNA from P_{RM} (Figure 17.22, part 4b). At the same time, transcription of the right early operon genes from P_R is decreased, but enough Q proteins are accumulated to set the genetic switch for transcription of the late genes for starting the lytic pathway (Figure 17.22, part 5b).

In sum, lambda uses complex regulatory systems to choose either the lytic or the lysogenic pathway. The decision depends on a sophisticated genetic switch that involves competition between the products of the *cI* gene (the repressor) and the *cro* gene (the gene for the Cro protein). If the repressor dominates, the lysogenic pathway is followed; if the Cro protein dominates, the lytic pathway is followed.

Summary

- In the lactose utilization system of *E. coli,* the addition of lactose to cells brings about a rapid synthesis of three enzymes. In the absence of lactose, the synthesis of the three enzymes is turned off. The genes for the enzymes are contiguous on the *E. coli* chromosome and are adjacent to two regulatory sequences: a promoter and an operator. The promoter, operator, and genes constitute an operon. Transcription of the genes results in a single polycistronic mRNA. A regulatory gene is associated with an operon. To turn on gene expression in the lactose system, a lactose metabolite binds with a repressor protein (the product of the regulatory gene), inactivating it and preventing it from binding to the operator. As a result, RNA polymerase can bind to the promoter and transcribe the three genes as a single polycistronic mRNA. Operons are commonly involved in the regulation of gene expression in a large number of prokaryotic and bacteriophage systems.
- If both glucose and lactose are present in the medium, the lactose operon is not induced, because glucose (which requires less energy to metabolize than does lactose) is the preferred energy source. This phenomenon is called catabolite repression and involves cellular levels of cyclic AMP. That is, in the presence of lactose and in the absence of glucose, cAMP complexes with CAP to form a positive regulator needed for RNA polymerase to bind to the promoter efficiently. The addition of glucose results in a lowering of cAMP concentration, so no CAP–cAMP complex is produced, and therefore RNA polymerase cannot bind efficiently to the promoter transcribe the *lac* genes.
- The expression of a number of bacterial amino acid synthesis operons is controlled by a repressor–operator system and through attenuation at a second regulatory sequence, called an attenuator. The repressor–operator system functions essentially like that for the *lac* operon, except that the addition of amino acid to the cell activates the repressor, thereby turning off the operon. An attenuator located between the operator region and the first structural gene is a transcription termination site that modulates the proportion of RNA polymerases that continue transcription past that point based on the level of tryptophan in the cell. Attenuation requires a tight coupling between transcription and translation, and the formation of particular RNA secondary structures that signal whether transcription can continue.
- The *ara* operon of *E. coli* encodes the genes for arabinose utilization. This operon has a regulatory gene, *araC*, the product of which (AraC) functions as a repressor and as an activator of transcription depending on the conditions. When arabinose is absent from the growth medium, AraC binds to regulatory sites, causing the DNA to loop and thereby blocking access of RNA polymerase to the structural gene's promoter. In this way, transcription of the structural genes is blocked by repression. When arabinose is present in the growth medium and glucose is absent, the operon is induced. Acting as an inducer, arabinose binds to AraC, changing its conformation. In its new form, AraC changes how it binds to DNA. No longer is a DNA loop formed, and RNA polymerase binding to the structural gene's promoter is facilitated by AraC functioning as an activator. In this sugar utilization system, catabolite repression operates as it does for the *lac* operon so the operon cannot be induced if glucose is present.
- Bacteriophages such as lambda are especially adapted for undergoing reproduction within a bacterial host. Many genes related to the production of progeny phages or to the establishment or reversal of lysogeny in temperate phages are organized into operons. Like bacterial operons, these operons are controlled through the interaction of regulatory proteins with operators and promoters that are adjacent to clusters of structural genes. Phage lambda has been an excellent model for studying the genetic switch that controls the choice between lytic and lysogenic pathways in a temperate phage.